Analytical Hierarchy Process Model for Selecting Road Contractors: A Case of Ethiopian Roads Administration

Eden Mershaye[1, *] and Abebe Dinku[1]

[1] *School of Civil and Environmental Engineering, College of Technology and Built Environment, Addis Ababa University, Addis Ababa, Ethiopia*

[*] *Corresponding Author's E-mail address: edenmershaye@gmail.com*

DOI: https://doi.org/10.63990/zede.v43i.12968

ABSTRACT

In the construction industry, the successful completion of a project mainly depends on the performance and capability of the contractor. It is therefore understood that the success of a project may be compromised without an appropriate and reliable process for identifying the most suitable contractor. Commonly, many organizations select contractors primarily on the lowest bid offer, as it is simple and emphasizes cost efficiency. However, this approach often undermines other important factors. Hence, incorporating a multi criteria decision making approach ensures a more balanced evaluation that considers both bid price and other key criteria. This study explores the application of the Analytical Hierarchy Process as a decision-making model for contractor's selection. A combination of descriptive and case study research design was employed to identify and prioritize the contractor selection criteria. An extensive literature review identified 23 contractor selection criteria, which were categorized into four main groups. A structured questionnaire was then used to collect primary data based on these criteria. Once the data is gathered, the criteria are ranked in order of significance and prioritized for use in the analytical hierarchy process model. The analysis revealed that past performance competency had the highest weight (38.56%) with a consistency ratio of 0.088, followed by technical competency (24.16%) and financial competency (20.42%). Management competency ranked lowest at 16.84%, with a consistency ratio of 0.0529. In summary, the proposed model enhances the contractor selection process by integrating multiple key competencies to evaluate contractors across broader criteria.

Keywords: Analytical Hierarchy Process, Bid Evaluation, Contractor selection, Multi criteria decision making.

1. INTODUCTION

The construction industry is a fundamental economic sector that influences the majority of other sectors. However, reports from 2021/2022 indicate that Ethiopia's construction industry has been experiencing fluctuations in growth due to economic and political changes [1]. From several subsectors, construction is one of the industries that contribute to the country's economy. Its major role in industrial production was indicated by its 72.2% share in the industrial sector and a 7.1% GDP growth rate in 2019/2020 G.C. [2]. Meanwhile, mining and quarrying, along with electricity and water, contributed 1.6% and 2.9% of the industrial sector, respectively. The share in industry of the construction sector and other industry sectors is typically shown in Figure 1 [3]. One of the most important public projects is construction of roads which act as a crucial infrastructure for the economic development of the country and helps to connect to other forms of transportation including railways, ships, and airplanes. For the fiscal year 2023-2024 the Ethiopian government has allocated 68.4 billion ETB to the Ethiopian Road Administration, comprising about 12 % of the country's total annual budget [4]. In a 2023 internal report by ERA, the administration noted that it oversees and

funds numerous road construction projects, with 231 being handled by various contractors [5]. Despite the economic significance of the road construction industry, some projects continue to suffer from delays and budget overruns. Evaluations conducted by project management and review committees have shown that contractors perform below expectations, primarily due to challenges in meeting contractual obligations within the planned cost and schedule parameters.

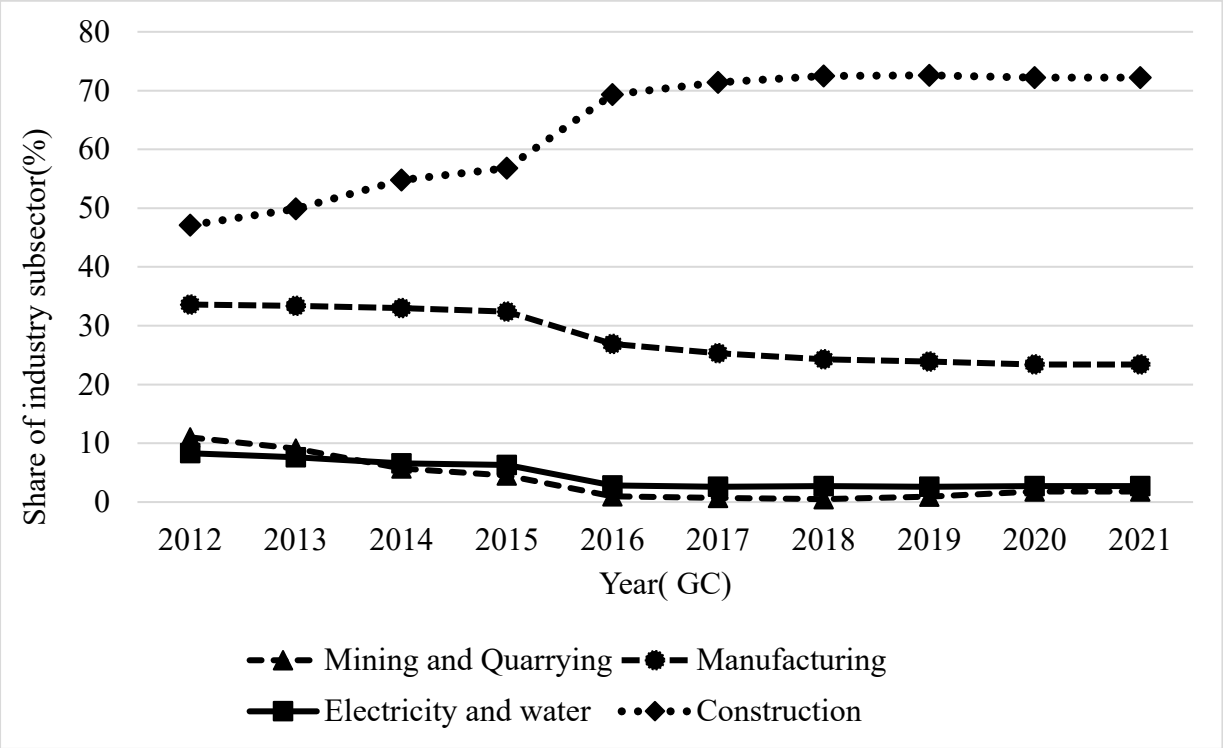


Figure 1: Construction sector's contribution to Ethiopia's domestic gross development [3]

It has been widely understood that, careful risk management is essential throughout all phases of a construction project from planning and design to execution and handover since each phase involves uncertainties related to cost, time, and quality. In this context, risk management refers to the systematic process of identifying, assessing, and mitigating potential risks that may influence project objectives. During the project delivery system, one of the most critical decisions during the procurement phase is contractor selection, which involves evaluating and awarding the contract to the most competitive and capable bidder. Given the complexity and inherent risks of construction activities, selecting the right contractor plays a decisive role in ensuring project success. Consequently, choosing the best contractor from the bids submitted to the client remains one of the most challenging tasks faced by clients in the construction industry [6]. The choice of contractor, among other factors, plays a significant role in determining the implementation and success of construction projects. This implies that low quality, schedule overruns, and low-cost performance will be experienced by the project if an unsuitable contractor is selected, [7], [8].

In Nigeria, open tendering in which the lowest bidder is awarded the contract, has become the most widely used method of contractor selection [6]. However, this method has been criticized by many researchers because contractors often submit artificially low bids during periods

of low activity, hoping to recover costs through subsequent claims [9]. To address such issues, computational models have been proposed to evaluate contractors based on project specifications and factual applicant data, thereby reducing the influence of opportunistic bidding. Consequently, contractor selection should consider multiple factors beyond cost alone. Most existing models, however, follow a relatively straightforward decision making procedure, without fully accounting for the chaotic and varied nature of contractor selection in a multi-criteria decision-making context [10].

Among the various Multi Criteria Decision Making methods available, the Analytic Hierarchy Process (AHP), developed by Saaty in the 1980s, is one of the most widely used techniques for contractor evaluation and selection [11]. It is a theory of measurement that is developed by pairwise comparison, and the priority scales are created using expert judgement. This study uses the Ethiopian Roads Administration (ERA) as a case study, an organization responsible for the construction of motorways, new link roads, rural and Woreda roads, rural and urban road rehabilitation and upgrading, and federal and regional road maintenance in conjunction with regional road authorities to meet the sector's goals. Further, the administration also goes through the process of selecting contractors in order to construct various road projects. In ERA, construction contracts are typically awarded based on the lowest bid offer, where the contract is granted to the firm that submits the lowest bid and meets the minimum qualification requirements. This approach is mandated under Ethiopia’s Public Procurement and Property Administration Proclamation No. 649/2009 and subsequent regulations, that emphasize transparency, fairness, and cost effectiveness in public procurements. Although cost focused selection ensures financial efficiency, it may not adequately reflect a contractor’s technical competence, experience, or ability to complete projects on time and within budget, potentially compromising project outcomes [12]. It has been argued that a common limitation of the low-bid approach is that contractors may submit unrealistically low bids, either intentionally or unintentionally, to secure the contract, which can lead to delays, quality issues, or disputes [12] . In contrast, several countries adopt multi-criteria evaluation methods, considering both cost and technical capabilities. For example, in the USA and many European countries, contractor selection often employes weighted scoring method, considering experience, financial stability, and past performance, aligning with best practice principles of project management [13], [14].

To address this limitation, this paper proposes an Analytical Hierarchy process based contractor selection model that incorporates both financial and technical criteria, offering a structured and objective alternative to the low bid approach. Accordingly, the study aims to identify major contractor selection criteria, determine the relative significance of each criterion using AHP, and develop a theoretical model for optimal contractor selection that aligns with international best practices.

1.1. Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP) is a mathematical technique that is well known for its effectiveness in solving complex decision-making problems. When decision makers have numerous criteria to consider, the AHP approach helps them select the optimal choice [15]. This method incorporates both quantitative and qualitative components and aids in organizing difficult decision-making issues into a step-by-step decision model. It is based on the notions that interconnections between clusters are unidirectional across the decision levels of the hierarchy and that there are no connections between

clusters and elements. The AHP process works as follows: First, it identifies criteria and the appropriate sub-criteria, breaking down the decision into distinct aspects to be taken into account. It then performs pairwise comparisons of the components to provide a relative relevance scale. In order to compare the significance of each pair of items at each level of the hierarchy, experts in the relevant domains are asked to determine the relative value of each criterion with respect to those at the second level. At the second level and beyond, experts compare the significance of every pair of sub-criteria under the same criterion Figure 2 illustrates the hierarchical structure of Analytical Hierarchy Process [16].

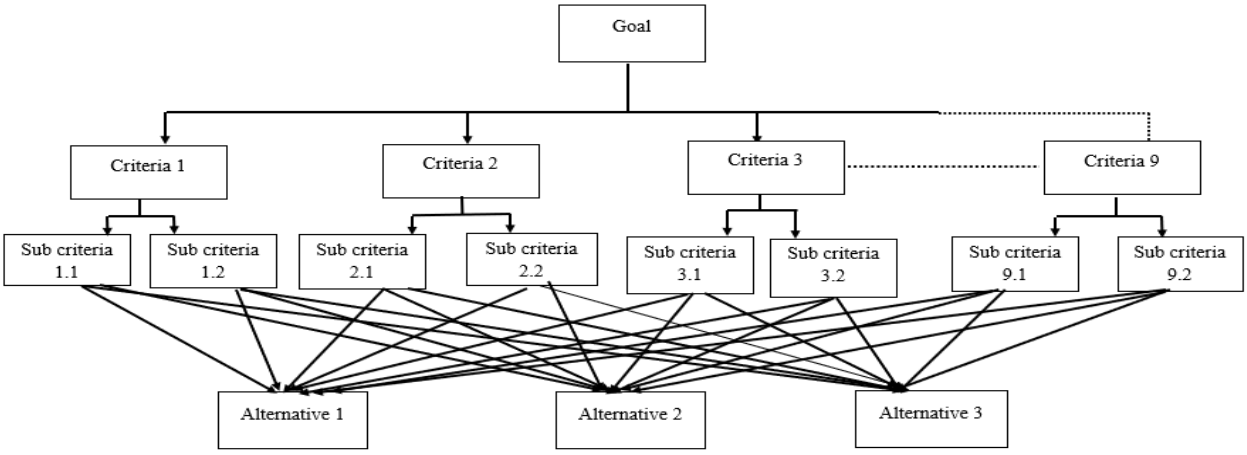


Figure 2: Typical hierarchical tree [16]

2. METHODS

Descriptive and case study research design is used for the study to obtain both a general understanding and an in-depth analysis of the contractor selection practices within the Ethiopian Roads Administration. The descriptive design helps to summarize and interpret the existing conditions, opinions, and practices related to contractor evaluation, while the case study approach enables a detailed examination of real-life projects to illustrate how the selection process is implemented in practice.

The study's target group are road contractors who were supervised by the Ethiopian Road Administration, professional engineers, project managers, road project consultants, tender assessment committees, and procurement specialists. Further, the study employs a sort of non-probability sampling method called purposive sampling to select contractors and consultants, and a probability sampling technique to select the employer. A purposive sampling method is used to select contractors and consultants because they possess relevant experience and specialized knowledge about the selection and evaluation process. On the other hand, a probability sampling technique is applied to select employers to ensure fair representation and reduce selection bias. Primary and secondary data sources were used for the analysis including; questionnaires, bid documents, financial data reports and published articles. Data presentation, frequency determination, and coding are done using SPSS Version 21.0.1.0. The Analytical Hierarchy Process (AHP) analysis in this study was

conducted using Super Decisions software 3.2.0. Figure 3 provides a summary of the total study methodology that is used in the current study.

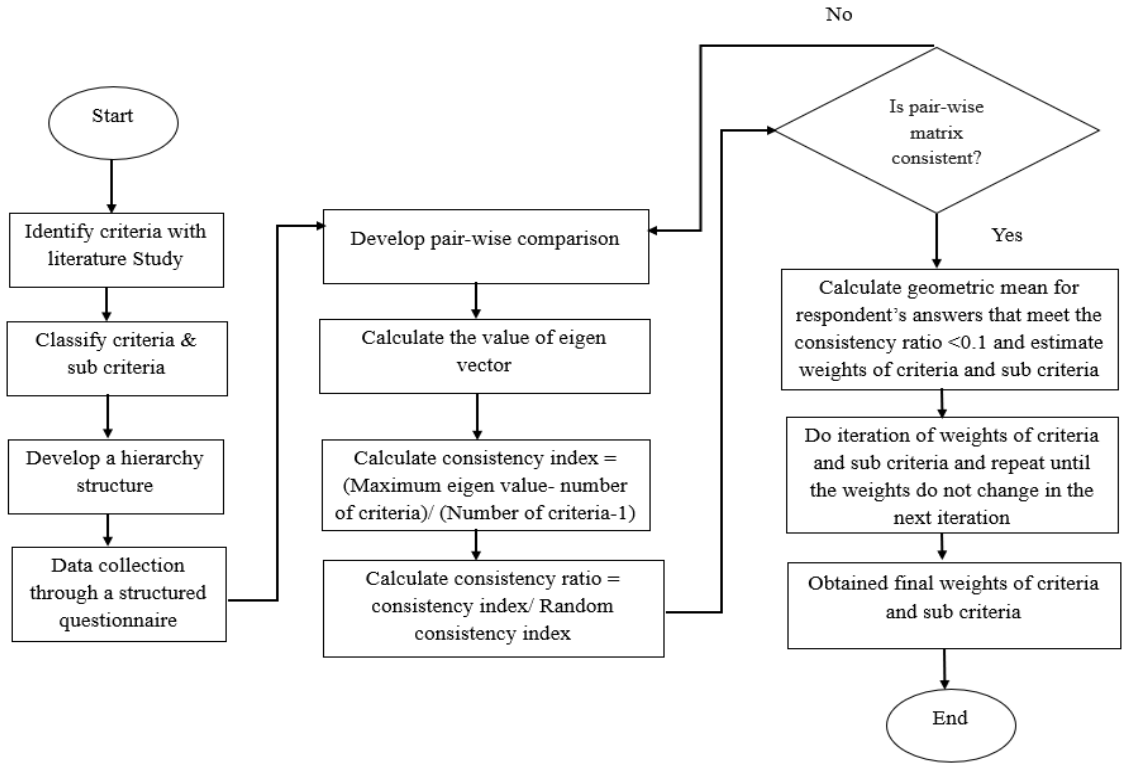


Figure 3: Research methodology

3. RESULTS AND DISCUSSIONS

3.1. Identification of Major Contractor Selection Criteria

The first objective of the study is to identify major contractor selection criteria. The study uses a mean score and ranking approach, using Equation 1. Table 1 shows that tender evaluation criteria answered by 87 respondents.

$$RII = \frac{5_{n_5} + 4_{n_4} + 3_{n_3} + 2_{n_2} + 1_{n_1}}{A * N} \quad (1)$$

Where:

RII= Relative Important Index

n_5= Number of respondents for strongly agree

n_4= Number of respondents for agree

$n_{3=}$ Number of respondents for neutral

$n_{2=}$ Number of respondents for disagree

$n_{1=}$ Number of respondents for strongly disagree

A= Highest weight

N= Total number of respondents

3.2. AHP Model Development

To implement the proposed AHP model effectively, a series of systematic steps must be followed. These procedures are essential for ensuring the accurate application of AHP in determining the contractor evaluation criteria. Below are the key steps that are applied:

Step 1: Determine the main criteria and sub-criteria that will be applied to the model (see Table 1).

Step 2: Take into consideration the most basic AHP scale for determining relative weights.

Step 3: The weights assigned to each criterion have been determined, and the matrix is created by averaging the responses from the three groups of participants who filled out the questionnaire. Here below is a pairwise comparison matrix that is used to assess a set of n criteria based on their respective weights. Equation 2 shows that a real matrix of size $m * m$ where m is the number of evaluation criteria taken in to account. The significance of each a_{ij} in matrix A represents, the j[th] criterion relative importance to the i[th] criterion.

$$A = a_{ij} = \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \dots & \cdots & \dots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \quad (2)$$

Here, the criteria in this case are $a_1, a_2 \cdots a_n$.Where"n" is the total number of criteria. A scale from 1 to 9 was used to establish the relative importance of the two criteria.

Table 1 Major contractors' selection criteria

Item	Criteria	RII	Mean	Rank
1	Annual turnover	0.9684	4.8419	1
2	Contractors working capital	0.9485	4.7425	2
3	Financial stability	0.9099	4.5493	3
4	Bid price	0.8897	4.4484	4
5	Cash flow projection	0.8617	4.3085	5
6	Expertise in the field	0.8171	4.0853	6
7	Not able to finish a contract	0.7926	3.9628	7
8	Past record conflict and dispute	0.7851	3.9253	8
9	The quantity of comparable work completed by the contractor	0.7776	3.8878	9
10	Performance of contractor on previously completed project	0.7662	3.8311	10
11	The quality of the project delivered through the allocated time and budget	0.7545	3.7725	11
12	Sufficient equipment/plant	0.7355	3.6774	12
13	Technical competency of staff members	0.7268	3.634	13
14	Contractors' knowledge regarding work methodology	0.7205	3.6024	14
15	Health and safety principles implementation	0.711	3.5549	15
16	The extent of QA/QC programs that have been used in previous projects	0.7045	3.5224	16
17	The comprehensiveness of the work technique description	0.6464	3.2321	17
18	Responding to tender specific requirements	0.6248	3.124	18
19	Project management skill	0.6026	3.013	19
20	Risk management	0.5951	2.9756	20
21	Number of personnel for the key position	0.5876	2.9381	21
22	Managerial capability	0.5242	2.6208	22
23	Subcontractors' management	0.4764	2.3819	23

3.3. AHP Process

As previously stated, there are four primary steps to implement AHP: i) structure the hierarchy model; ii) computing pairwise comparisons; iii) synthesizing and computing decisions for weight prioritization; and iv) assessing the consistency of the outcomes.

3.3.1. Structure of the hierarchy

By using a pairwise comparison between each data set's criteria, the AHP divides the contractor selection process into a hierarchy consisting of the Goal, Criteria, and Alternatives. Financial, past performance, technical, and management competency are the factors that AHP uses to conduct the pairwise comparison. The alternatives are the number of contractors bidding for the contract. Table 2 represents Saaty rating scale which is used to determine the relative weights for the selection criteria.

Table 2 The Saaty rating scale [17].

Intensity of importance	Definition	Explanation
1	Equal importance	i and j are equivalent
3	Somewhat more important	i is slightly preferred to j
5	Much more important	i is strongly preferred to j
7	Very much important	i is very strongly preferred to$j$.
9	Absolutely more important	i is absolutely preferred to j
2,4,6,8	Intermediate values	When compromised is needed

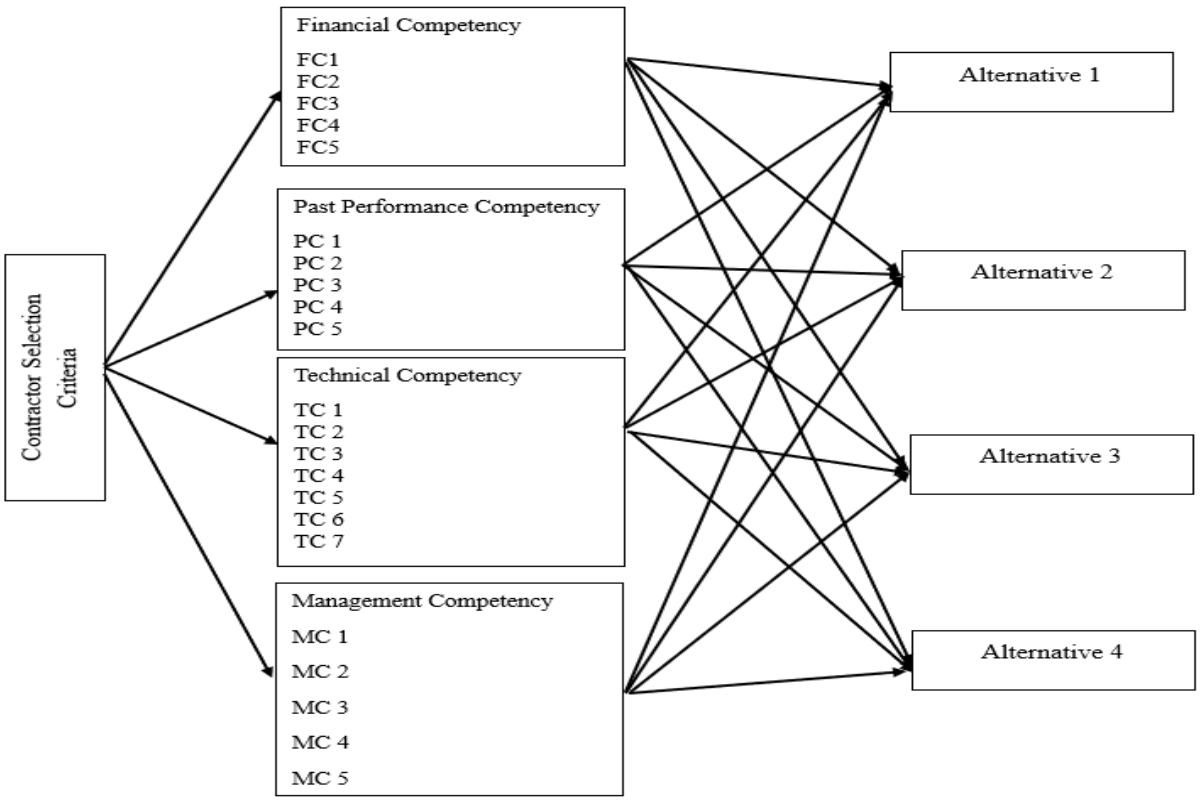


Figure 4 A hierarchical representation of the AHP model.

3.3.2. Performing Pair Wise Comparisons

After developing the hierarchy, pairwise comparison was methodically carried out across the structure. Table 3 presents the summary of results of the pairwise comparison of the criteria for each of the four primary criteria: Financial competency (FC), Past performance competency (PC), Technical Competency (TC), and Management competency (MC).

Table 3 Pair wise comparison matrix

Criteria	FC 1	...	FC 5	Eigenvector	Criteria vector (w)
FC 1	1	...	FC_{15}	$\left(v_{ij}=\prod_{k=1}^{k} a_{ij}\right)^{\frac{1}{k}}$	$Wi=\frac{vi}{\sum vi}$
⋮	⋮	1	⋮		$Wi=\frac{vi}{\sum vi}$
FC 5	$1/_{FC_{15}}$	...	1		$Wi=\frac{vi}{\sum vi}$
Eigenvalue λmax	$\sum c_{ji*w_i}$				$\sum w_{i=1}$
Consistency Ratio	$\frac{\frac{\lambda max-n}{n-1}}{RI}$				

Criteria	PC 1	...	PC 6	Eigenvector	Criteria vector (w)
PC 1	1	...	PC_{16}	$\left(v_{ij}=\prod_{k=1}^{k} a_{ij}\right)^{\frac{1}{k}}$	$Wi=\frac{vi}{\sum vi}$
⋮	⋮	1	⋮		$Wi=\frac{vi}{\sum vi}$
PC 6	$1/_{PC_{16}}$	...	1		$Wi=\frac{vi}{\sum vi}$
Eigenvalue λmax	$\sum c_{ji*w_i}$				$\sum w_{i=1}$
Consistency Ratio	$\frac{\frac{\lambda max-n}{n-1}}{RI}$				

Criteria	TC 1	...	TC 6	Eigenvector	Criteria vector (w)
TC 1	1	...	TC_{16}	$\left(v_{ij}=\prod_{k=1}^{k} a_{ij}\right)^{\frac{1}{k}}$	$Wi=\frac{vi}{\sum vi}$
⋮	⋮	1	⋮		$Wi=\frac{vi}{\sum vi}$
TC 6	$1/_{TC_{16}}$	...	1		$Wi=\frac{vi}{\sum vi}$
Eigenvalue λmax	$\sum c_{ji*w_i}$				$\sum w_{i=1}$
Consistency Ratio	$\frac{\frac{\lambda max-n}{n-1}}{RI}$				

Criteria	MC 1	...	MC 5	Eigenvector	Criteria vector (w)
MC 1	1	...	MC_{15}	$\left(v_{ij}=\prod_{k=1}^{k} a_{ij}\right)^{\frac{1}{k}}$	$Wi=\frac{vi}{\sum vi}$
⋮	⋮	1	⋮		$Wi=\frac{vi}{\sum vi}$
MC 5	$1/_{MC_{15}}$	...	1		$Wi=\frac{vi}{\sum vi}$
Eigenvalue λmax	$\sum c_{ji*w_i}$				$\sum w_{i=1}$
Consistency Ratio	$\frac{\frac{\lambda max-n}{n-1}}{RI}$				

Where: FC= Financial competency, PC= Past perormance competency, MC= Management competency, TC= Technical competency, λmax= Maximun eigen value, n = Number of criteria w = Criteria vector, RI = Random consistency index, $a_{ij=}$ pairwise compasrsion values.

3.3.3 Weight prioritization and consistency assessment in decision making

After structuring the hierarchy and forming the pairwise comparisons, the next step is calculation of weights and consistency checking (Table 4).

3.4. Discussion of AHP Results

The current study and ERA's approach consider similar evaluation elements, such as financial competency, technical competency, past performance, available resource (equipment, personnel), and legal compliance. For example, ERA includes requirements like financial performance history, experience, and key personnel availability, which aligns with the sub criteria that is proposed in the current research. Further, the current study highlighted past performance competency as the most important major category, particularly noting factors like project quality and past conflict history. ERA also evaluates contractors' non- performance history and pending litigation, as a part of its post qualification examination ensuring that contractors with the history of non-performance are filtered out.

Table 4 Summary of AHP results for main and sub criteria for selection of contractor

Criteria	Aggregated pairwise matrix $\left(v_{ij}=\prod_{k=1}^{k} a_{ij}\right)^{\frac{1}{k}}$	Sub criteria	Criteria weight $(Wi = vi/\sum vi)$	Criteria (%)	Rank	Consistency
Main criteria (FC, PC, TC & MC)	$\begin{bmatrix} 1 & 1/2 & 1/2 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 1/2 & 1 & 1 \\ 1/2 & 1/2 & 1 & 1 \end{bmatrix}$	FC	0.2042	20.66%	3	$\lambda max = 4.19$
		PC	0.3856	38.54%	1	$CI = 0.06$
		TC	0.2416	23.81%	2	$CR = 0.0694$
		MC	0.1684	16.99%	4	
FC (Financial competency)	$\begin{bmatrix} 1 & 1/3 & 1 & 1/2 & 1/2 \\ 3 & 1 & 1 & 1/2 & 2 \\ 1 & 1 & 1 & 1 & 3 \\ 2 & 2 & 1 & 1 & 2 \\ 2 & 1/2 & 1/3 & 1/2 & 1 \end{bmatrix}$	FC 1	0.1201	12.25%	5	$\lambda max = 5.333$
		FC 2	0.2314	22.98%	3	
		FC 3	0.2373	23.75%	2	$CI = 0.083$
		FC 4	0.2784	27.76%	1	$CR = 0.07368$
		FC 5	0.1326	13.26%	4	
PC (Past performance competency)	$\begin{bmatrix} 1 & 1/3 & 1/4 & 4 & 1/2 & 1/2 \\ 3 & 1 & 1/2 & 3 & 1 & 1/3 \\ 4 & 2 & 1 & 4 & 1 & 1/3 \\ 1/4 & 1/3 & 1/4 & 1 & 1/3 & 1/3 \\ 2 & 1 & 1 & 3 & 1 & 1/3 \\ 2 & 3 & 3 & 3 & 3 & 1 \end{bmatrix}$	PC 1	0.1000	10.64%	5	$\lambda max = 6.566$
		PC 2	0.1485	14.82%	4	
		PC 3	0.2105	20.67%	3	$CI = 0.113$
		PC 4	0.0522	5.46%	6	$CR = 0.0880$
		PC 5	0.1493	14.85%	2	
		PC 6	0.3391	33.55%	1	
Criteria	Aggregated pairwise matrix $\left(v_{ij}=\prod_{k=1}^{k} a_{ij}\right)^{\frac{1}{k}}$	Sub criteria	Criteria weight $(Wi = vi/\sum vi)$	Criteria (%)	Rank	Consistency
TC (Technical competency)	$\begin{bmatrix} 1 & 1/5 & 1/4 & 1/3 & 1/2 & 1/3 \\ 5 & 1 & 1 & 1/3 & 3 & 3 \\ 4 & 1 & 1 & 1/3 & 1 & 1/2 \\ 3 & 3 & 3 & 1 & 4 & 4 \\ 2 & 1/3 & 1 & 1/4 & 1 & 1/3 \\ 3 & 1/3 & 2 & 1/4 & 3 & 1 \\ 4 & 3 & 3 & 2 & 3 & 2 \end{bmatrix}$	TC 1	0.0419	4.33%	7	$\lambda max = 7.647$
		TC 2	0.1526	15.25%	3	
		TC 3	0.0932	9.56%	5	$CI = 0.107$
		TC 4	0.2577	24.77%	2	$CR = 0.0794$
		TC 5	0.0636	6.71%	6	
		TC 6	0.1139	12.06%	4	
		TC 7	0.2769	27.31%	1	
MC (Management competency)	$\begin{bmatrix} 1 & 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 1 & 3 \\ 1/2 & 1/2 & 1 & 1/3 & 1/2 \\ 1 & 1 & 3 & 1 & 3 \\ 1/2 & 1/3 & 2 & 1/3 & 1 \end{bmatrix}$	MC 1	0.2382	23.88%	1	$\lambda max = 5.23$
		MC 2	0.2618	25.99%	3	$CI = 0.058$
		MC 3	0.0981	9.98%	5	$CR = 0.0529$
		MC 4	0.2810	27.99%	2	
		MC 5	0.1206	12.16%	4	

A primary difference lies in AHP's weighted scoring VS ERA's pass/fail approach. While the AHP analysis assigns weights to prioritize criteria, ERA does not apply weightage or scoring; contractors only need to meet the minimum

requirements. In ERA system, all criteria are essential but do not contribute differently to an overall score, where as in the current study criteria are weighted. In addition, ERA's financial evaluation phase solely determines the contract award by selecting the lowest bid, assuming all technical requirements are met. In contrast, AHP approach ranks criteria, which support a best-value selection method rather than strictly selecting the lowest bid.

To ensure the reliability and validity of the results, the research instruments underwent both reliability testing and content validity assessment. The reliability analysis, performed using Cronbach's Alpha (α), showed acceptable to excellent internal consistency across all major variables: financial competency (α = 0.777), management competency (α = 0.748), technical competency (α = 0.935), and past performance competency (α = 0.912). A reliability value that gives the value of α > 0.9 (Excellent); α > 0.8 (Good); α > 0.7(Acceptable); α > 0.6 (Questionable); α > 0.5 (Poor) and α< 0.5 (Unacceptable) [18]. The study therefore used the Cronbach's value that ranged between 0.7 and above to determine reliability of instruments. In addition, content validity was ensured through a pilot study involving construction industry professionals, followed by a review by colleagues with experience in the road construction sector. Feedbacks from the pilot test helped refine the questionnaire and confirm that each item adequately represented the intended construct.

Therefore, the findings presented in this study are based on validated and reliable data, ensuring that the comparison between the current study and ERA's approach is both credible and methodologically sound. If ERA aims to advance its contractor selection approach beyond minimum standards, adopting a weighted criteria method such as AHP could help prioritize contractors who not only offer cost effectiveness but also demonstrate higher competency essential for project success.

3.5. Case Study

A case study was conducted on a tender that was announced in the *Ethiopian Herald* newspaper and on the Ethiopian Roads Administration's website on September 29 and 30, 2020, for the road construction project. This particular tender was selected because it represents a typical ERA procurement process, with clearly defined prequalification and bid evaluation stages that align with the objectives of this research. In addition, the Ethiopian Road Administration provide complete and accessible documentation, including bid evaluation reports, financial and technical data, which made it suitable for applying both the existing ERA selection procedure and the proposed AHP based model.

The employer sought to assess potential bidders capable of successfully completing the project. Out of nine bidders, four contractors passed the prequalification round and were included in the analysis. These four contractors were first evaluated using the ERA's existing selection approach, and then re-assessed using the proposed AHP-based model to compare outcomes. Hence, the scenario described in the case study represents a shortlist of four contractors considered for contract award.

3.5.1. Contractor's profile

The contractor's profile is studied thoroughly, and the vast details about their turnover, projects completed by them, equipment, machines, and tools owned or required by them, and the technical personnel available with them are consolidated to create a short profile that gives an idea of their strengths and weaknesses in the respective criteria under consideration. The contractor's profile is shown in Table 5 in tabular form.

Two processes were used in the evaluation of bids: post-qualification and financial evaluation. After assessing the post-

qualification applications in accordance with the requirements specified in the bidding document, the Tender Assessment Committee advised all bidders who had submitted both their financial bids and qualification documents to proceed with the opening of their financial bids. Based on the outcome of the evaluation, the Tender Assessment Committee recommended that Contractor 2 be invited for pre-contract discussions and the eventual award of the construction works.

Table 5: Contractor's profile

Item no.	Applicants	Grade	Country of Registration
1	Bidder 1	GC-1	China
2	Bidder 2	GC-1	
3	Bidder 3	GC-1	Ethiopia
4	Bidder 4	GC-1	

3.5.2. Contractor selection based on AHP model

Following a prequalification process that involved reviewing the contractors' files and performance records, only four contractors were found to be eligible for the project. This project, together with these four contractors, serves as a prototype for validating the proposed AHP based model. The decision makers evaluated each contractor's performance in relation to the twenty-three sub criteria outlined in the previous section, which represent the main criteria's that are used in contractor selection process.

The evaluation criteria were grouped under four major categories; past performance competency, technical competency, financial competency and management competency. The alternative options under consideration were designated as Bidder 1, Bidder 2, Bidder 3, and Bidder 4.

The first phase in the Analytical Hierarchy Process (AHP) involves clearly formulating the problem and establishing the decision hierarchy, which defines the goal, criteria, and sub-criteria before moving to the pairwise comparison stage. Figure 4 shows the hierarchical structure and the flow of decision making applied in this study. After defining the criteria, Saaty's fundamental scale of pairwise comparisons was used to assess the relative importance of one element over another. This scale allows decision makers to express judgments consistently and to quantify qualitative assessments, which is particularly useful in multi criteria decision making contexts such as contractor selection. Subsequently, for Bidder '1', '2', '3', and '4', their respective weights, pairwise comparison matrices, consistency index, and consistency ratio were computed to ensure logical consistency of the judgements. The results of these comparisons are presented in Table 6.

Table 6 Pairwise comparison and normalization matrices for each criterion

Pairwise Comparison Matrix for Financial Competency					**Normalization Matrix**					
FC	Bidder1	Bidder2	Bidder 3	Bidder 4	FC	Bidder 1	Bidder 2	Bidder 3	Bidder 4	Criteria weight
Bidder 1	1	$1/4$	$1/3$	$1/8$	Bidder 1	0.0625	0.0285	0.0357	0.0871	0.0534
Bidder 2	4	1	2	$1/7$	Bidder 2	0.2500	0.1142	0.2142	0.0995	0.1695
Bidder 3	3	1/2	1	$1/6$	Bidder 3	0.1875	0.0571	0.1071	0.1161	0.1169
Bidder 4	8	7	6	1	Bidder 4	0.500	0.800	0.6428	0.6970	0.6599
Sum	16	8.75	9.33	1.434	$\boldsymbol{\lambda max = 4.202, CI = 0.067\ RI = 09\ , CR = 0.074}$ < 0.1 Ok!					

Pairwise Comparison Matrix for Past Performance Competency

PC	Bidder 1	Bidder 2	Bidder 3	Bidder 4
Bidder 1	1	1/4	1/3	$1/7$
Bidder 2	4	1	3	$1/5$
Bidder 3	3	1/3	1	$1/6$
Bidder 4	7	5	6	1
Sum	15	6.583	10.33	1.509

Normalization Matrix

PC	Bidder 1	Bidder 2	Bidder 3	Bidder 4	Criteria weight
Bidder 1	0.0667	0.0379	0.0322	0.0946	0.0578
Bidder 2	0.2667	0.1518	0.2903	0.1324	0.2103
Bidder 3	0.200	0.0506	0.0967	0.1104	0.1144
Bidder 4	0.4667	0.7594	0.5806	0.6624	0.6173

$\boldsymbol{\lambda max = 4.228, CI = 0.076\ RI = 0.9\ , CR = 0.0844}$ < 0.1 Ok!

Pairwise Comparison Matrix for Technical Competency

TC	Bidder 1	Bidder 2	Bidder 3	Bidder 4
Bidder 1	1	$1/5$	$1/2$	$1/7$
Bidder 2	5	1	2	$1/6$
Bidder 3	2	$1/2$	1	$1/5$
Bidder 4	7	6	5	1
Sum	15	7.7	8.5	1.509

Normalization Matrix

TC	Bidder 1	Bidder 2	Bidder 3	Bidder 4	Criteria weight
Bidder 1	0.0667	0.0259	0.0588	0.0946	0.0615
Bidder 2	0.3333	0.1297	0.2352	0.1104	0.2022
Bidder 3	0.1333	0.0649	0.1176	0.1324	0.1121
Bidder 4	0.4666	0.7792	0.58823	0.6624	0.6241

$\boldsymbol{\lambda max = 4.202, CI = 0.067\ RI = 0.9\ , CR = 0.0750}$ < 0.1 Ok!

Pairwise Comparison Matrix for Management Competency

MC	Bidder 1	Bidder 2	Bidder 3	Bidder 4
Bidder 1	1	$1/4$	$1/3$	$1/6$
Bidder 2	4	1	2	$1/5$
Bidder 3	3	$1/2$	1	$1/5$
Bidder 4	6	5	5	1
Sum	14	6.75	8.33	1.566

Normalization Matrix

MC	Bidder 1	Bidder 2	Bidder 3	Bidder 4	Criteria weight
Bidder 1	0.0714	0.0370	0.040	0.1063	0.0637
Bidder 2	0.2857	0.1481	0.240	0.1276	0.2003
Bidder 3	0.2142	0.0740	0.120	0.1276	0.1340
Bidder 4	0.4285	0.7407	0.600	0.6382	0.6019

$\boldsymbol{\lambda max = 4.189, CI = 0.0630\ RI = 0.9\ , CR = 0.0701}$ < 0.1 Ok!

3.6. Summary of the Results on Case Study

As it can be observed from Table 6, Bidder 4 was determined to have the highest index based on the case study's findings. The comparison of bidders' ranks based on established models and current practice is shown in Table 7.

Table 7 Rank of bidders based on the two methods

Bidders	Rank based on current practice	Rank based on developed model
1	3	4
2	1	2
3	2	3
4	4	1

The differing results between the Ethiopian Road Administration method and the Analytical Hierarchy Process (AHP) can be justified by recognizing the broader scope and comprehensive evaluation offered by the AHP method. While the administration prioritizes cost efficiency, often resulting in selecting the lowest bidder, the AHP approach considers multiple criteria, such as technical expertise, financial stability, past performance, and project management capabilities. This broader evaluation ensures that the selected contractor is qualified to deliver high quality outcomes and minimize project risks. Hence, adopting the AHP technique introduces a structured and transparent decision-making framework that balances cost

considerations with quality and performance factors. It enhances consistency in evaluation, reduces subjectivity, and supports better-informed contractor selection decisions, which ultimately contribute to improved project outcomes.

3.7. Sensitivity Analysis

Super decision software version 3.2 was used to do the sensitivity analysis, which tested the adaptability and accuracy of multi criteria judgements through criteria change. To see how the weights of the main criteria; financial competency, past performance competency, technical competency and managerial competency affected the bidders' overall rating, sensitivity analysis was carried out, $\pm 10\%$ adjustment was made to each criterion. The sensitivity analysis conducted using super decision software demonstrates that despite varying the percentage of weight assigned to different criteria, there is no change in the contractor selection ranking. The analysis confirms that the contractor rankings remain consistent, and no rank reversal point is observed.

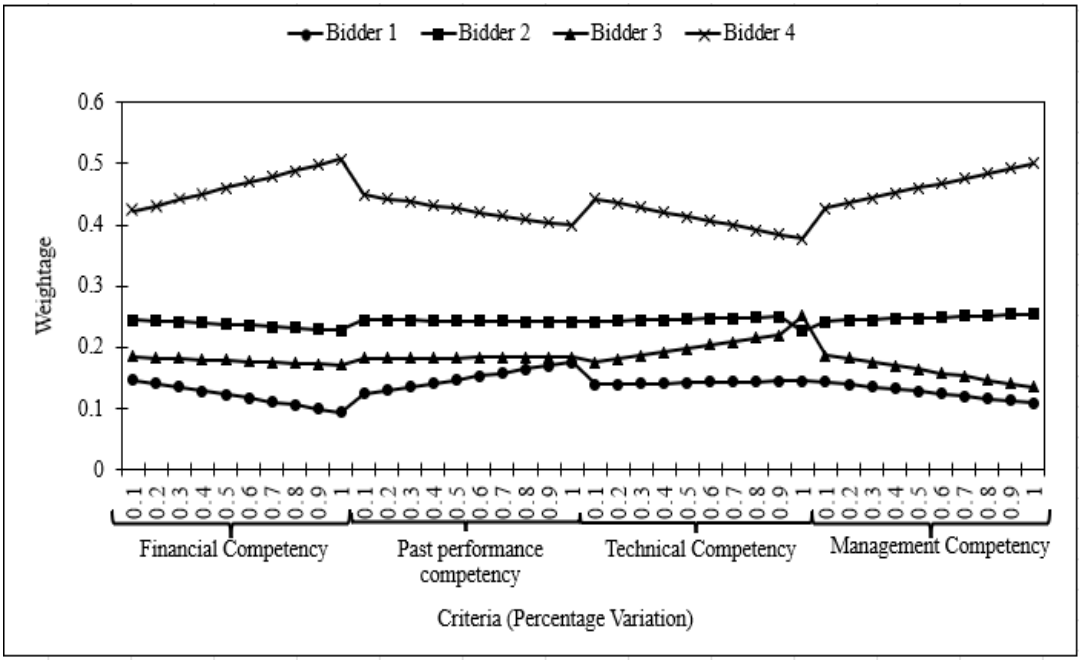


Figure 5 Sensitivity analysis for various percentage

4. CONCLUSIONS

A total of 23 contractor selection criteria were identified through an extensive literature review. These criteria were categorized into four main groups: financial, technical, past performance, and management competency. The Analytic Hierarchy Process (AHP) model assigned weights to these four main criteria to reflect their relative importance in the contractor selection process. The results from the AHP analysis showed that past performance competency was the most critical factor, carrying the largest weight of 38.56%. Following past performance, technical competency was the second most important criterion, with a weight of 24.16%. Technical competency includes factors like the availability of specialized equipment, the expertise of technical staff, and the contractor's ability to implement health, safety, and quality control measures.

Financial competency was ranked third, with a weight of 20.42%, this criterion evaluates a contractor's financial health, including their cash flow projections, bid price, and financial stability, annual turnover and working capital. Lastly, management competency ranked fourth,

with a weight of 16.84%. Although it is the least weighted of the four main criteria, it still plays an essential role in determining a contractor's ability to manage resources, personnel, and subcontractors effectively.

A case study was conducted to the proposed AHP model. The model was tested with four contractors engaged in an ERA road construction project. Based on the AHP assessment, Contractor 4 was selected as the most suitable choice. To ensure the scientific validity of the model, consistency tests were carried out. The AHP consistency ratio (CR) for all pairwise comparison matrices were found to be below the acceptable threshold of 0.10, confirming logical consistency in the judgments provided by respondents. However, since the validation was based on a single case study, future research is recommended to apply the model to multiple projects across different contexts and include additional selection criteria to enhance the model's generalizability and robustness.

The proposed AHP-based decision-making model enhances the contractor selection process by integrating multiple key competencies including; past performance, technical ability, financial stability, and management skills. This multi-criteria approach ensures that contractors are evaluated on a broader set of factors, which not only improves project outcomes but also encourages higher standards in the industry.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLWDGEMENT

The authors acknowledge the support for College of Technology and Built Environment.

REFERENCES

[1] H. Suliman, "Causes and Effects of Cost Overrun on Construction Project in Bahrain: Part I (Ranking of cost overrun factors and risk mapping)," *Modern Applied Science,* vol. 11, no. 7, 2017.

[2] Plan and Development Commision, "Annual Report, Ethiopia," 2019-2020.

[3] National Bank of Ethiopia, " Annual Report," 2021/2022.

[4] Ministry of Finance, "2016 FY GoE Budget summary," Addis Ababa, 2016.

[5] Ethiopia Roads Administration, "Construction Projects 2015 Implementation & 2016 E.C Plan," 2016.

[6] O. Olatunji, "Bid or no-bid decision factors of indigenous contractors in Nigeria," *Engineering, Construction and Architectural Management, 24(3),,* pp. 378-392, 2017.

[7] D. Ayettey, "Contractor selection Criteria in Ghanaian Construction Industry: Benefits and Challenges.," *Journal of Building Construction and Planning Research,6,* pp. 278-297, 2018.

[8] A. Khoso, "Embedded Remote Group Environment Through Modification In Macbeth- An Application of contractor's Selection In Construction.," *Journal of Civil Engineering and Management,27(8),* pp. 595-616, 2021.

[9] A. Birjandi, "Evaluation and selecting the contractor in bidding with incomplete information using MCGDM method," *Soft computing,* vol. 23, no. 20, pp. 10569-10585, 2019.

[10] S. Bishaw, "Development of Fuzzy Decision Model for Ethiopian Construction contractors selection," *PHD diss., Addis Ababa Science and Technology University,* 2019.

[11] M. Rahimdel, "Application of analytical hierarchy process to selection of primary crusher," *International Journal of Mining Science and Technology, 24(4),,* pp. 519-523, 2014.

[12] M. Demesew, "Assessment of Tender Evaluation System for Public Building Project Works in Ethiopia," *Thesis, Jimma University,* 2019.

[13] M. Hasnain, "Best Value Contractor Selection in Road Construction Projects: ANP- Based Decision Support System," *International Journal of Civil Engineering,* vol. 16, pp. 695-714, 2017.

[14] Y. Khoso, "'Extended review on contractor selection in construction projects'," *Canadian Journal of Civil Engineering, Vol. 47, No. 7,* pp. 771-789, 2020.

[15] C. Benfares, "Multi-Criteria Decision Making Semantic for Mental Healthcare. In Research Anthology on Mental Health Stigma, Education, and Treatment," *IGI Global: Hershey, PA, USA,* pp. 178-192, 2021.

[16] S. Anna, "Analytic Hierarchy Process in Multiple–Criteria Decision–Making: A Model Example," *SHS Web of Conference 90,0109,* pp. 1-10, 2021.

[17] R.W. Saaty, "The analystic Hierarchy Process_What it is and How it is used," *Pergamon Journals Ltd,* vol. 9, pp. 161-176, 1987.

[18] C. Sylvester, "Cronbach's Alpha: A Cornerstone in Ensuring Reliability and Validity in Environental Health Assessment," *ES Energy & Environment,* pp. 1-14, 2023.

Safety Evaluation of Reinforced Concrete Highway Bridges Under Overloaded Truck Traffic: A Data-Driven Assessment Using Static Weighing Station Records

Yisihak Gebre Tarekegn [1,*], Biniam Wessenseged [2], Abrham Gebre [1], Getu Segni [1] and Asmerom Weldegerima[1]

[1] *School of Civil and Environmental Engineering, College of Technology and Built Environment, Addis Ababa University, Addis Ababa, Ethiopia*

[2]*Save the Children International, Addis Ababa, Ethiopia*

**Corresponding Author's E-mail address: yisihak.gebre@aau.edu.et*

DOI: https://doi.org/10.63990/zede.v43i.12969

ABSTRACT

Assessing performance assessment of reinforced concrete (RC) highway bridges subjected to overloaded truck is important in maintaining safety and sustainability of transport infrastructures. These trucks cause threat to bridges and lead to deterioration if not managed. The framework of assessment consists of structural analysis techniques, load rating methodologies, condition assessment procedures, and risk evaluation criteria. Studies showed that bridges in Ethiopia are overloaded and hence, in this study, a comprehensive safety assessment of selected RC highway bridges subjected to overloaded truck is presented. Nine RC girder bridges found along the selected routes have been considered for investigation. To investigate the effects of overloaded vehicles on Ethiopian bridges, 51,900 actual truck loading data from three static weighing stations (SWS) were collected over a period of five years. Rating factors for bridges were determined based on legal loads, actual truck load data, and extrapolated load data, taking into account the estimated remaining service life of the bridges and possible future reinforcement corrosion. The results revealed that, on average, 16.3 % and 33.85 % of the trucks violated the limit set on national regulation and bridge formulas, respectively. In addition, the rating factors for the bridges were reduced by 30.18 % and 56.29 % for the actual truck load data and extrapolated load data, respectively, compared to the legal loads. The result showed the bridges' performance is severely affected and hence enforcing the current law and developing appropriate mitigation strategies are recommended.

Keywords: Bridges, Overloaded trucks, RC, Rating factors, Safety, SWS.

1. INTRODUCTION

1.1 Overloaded Trucks

Overloaded trucks on bridge structures create difficulty due to the growing business activity and rising need for transportation infrastructure. Moreover, the susceptibilities of RC highway bridges to truck traffic are made worse by elements like aged infrastructure, poor maintenance methods, and changing regulatory standards. Engineers and researchers have created thorough methods and strategies for evaluating the performance of bridges under truck loads that are too heavy in response to these issues. These procedures include complex load rating techniques [1], sophisticated structural analysis tools, condition assessment techniques [2], risk evaluation criteria, and mitigation measures that are specifically designed to meet the requirements of highway bridges made of RC [3], [4].

The Ethiopian Roads Administration (ERA) has employed a rating legal truck model for the structural evaluation of highway bridges [5] which was adopted

from [6]. This model can be used as a standard load for the safety evaluation of bridges in general but fails to represent the actual loading conditions of the country as there is a significant change in truck loads [7]. This has significant impacts on the safety, reliability, and maintenance of bridge structures and systems. Hence, a site-specific live load model which accounts the loading conditions of the country must be employed. The reliability of RC highway bridge is extremely important for uninterrupted and safe traffic flow. In Ethiopia, however, the increasing tendency to overloaded vehicles poses threats to these critical structures. These trucks not only put excess weight on bridges but also lead to potential failures.

Safety evaluation of RC girder bridges for the overloaded vehicles is paramount to ensure the structure's integrity and safety to the public. Using the methods of visual inspections, non-destructive tests (NDT), load testing and continuous monitoring, one can predict and prevent the risks associated with loading in structural systems. Applying design improvements, appropriate overload management techniques, and strict maintenance procedures can greatly enhance the durability and reliability of such structures [8]-[10]. Timely safety evaluation of bridges reduces rehabilitation costs and avoids unnecessary replacement of highway structures that are still in serviceable condition [11]. Several authors have paid much attention on research on the impact of overloaded trucks on performance of bridges. Increased stresses from overloads and accelerated material degradation caused by corrosion reduce the service life of bridges and pose a threat to safety and operational performance [12], [13].

Evaluation of existing bridges in Texas built in the 1950s and 1960s was conducted using the LRFR (Load and Resistance Factor Rating) method recommended by AASHTO (American Association of State Highway and Transportation Officials). The findings revealed that the calculated rating factors are lower than those required for the minimum design vehicle specified in the Manual for Condition Evaluation of Bridges (MCEB) [14]. In a similar study conducted in New York state, bridge load ratings were performed, revealing that the LRFR legal load rating factors were 53% higher than the inventory legal load factors determined using the LFR (Load Factor Rating) method [15].

1.2 Data Extrapolation

In some cases, heavy trucks may have bypassed the weighing stations and critical data may not be recorded. To account these missed data, extrapolation of the recorded extreme vehicle loadings is usually done, commonly, the maximum load effects which could occur within a 75-year return period [16]. There are basically several scenarios for selecting population size [17], [18]. Among them, the block maxima method grouped the data on a monthly basis and one extreme data is selected from each month [19]. Castillo's approach uses the top $2\sqrt{n}$ data [20], and Enright's fitting method selects the top 30% data [7], [21].

In extrapolating the traffic data, in this study, a Sivakumar extrapolation approach is selected over other extrapolation methods as the method was developed specifically for highway bridges to estimate maximum live load effects (like bending moment or shear force) for long return periods (e.g., 75 or 100 years) [22] and its practical suitability for estimating extreme bridge load effects using limited high-end traffic data [23]. Moreover, it provides a simplified empirical model based on observed traffic load effects, extrapolating rare extreme events through a fitted curve with reasonable accuracy, and has been successfully applied in the load rating of bridge structures [22]. Whereas generalized extreme value-based methods and block maxima techniques

typically require extensive, continuous data and rely on specific distribution assumptions [23], [24]. Similarly, Enright [18] used data extrapolation by fitting a Weibull extreme value distribution to extrapolate bridge load effects which involve reliability-based formulations, making them data- and computation-intensive.

Hence, in the present study, extrapolated bending moment and shear force for a return period of 75 years were calculated using Eqn. (1) [25].

$$\mu_{max} = \mu + \sigma\sqrt{2\ln(N)} - b$$
$$b = \sigma\frac{\ln[\ln(N)] + \ln(4\pi)}{2\sqrt{2\ln(N)}} \quad (1)$$

Where N is the number of data in a specific return period, μ_{max} is the extrapolated mean value and σ is the standard deviation.

1.3 Reinforcement Corrosion

In addition to increasing load intensity, corrosion of reinforcing steel reduces cross-sectional area, which affects load-carrying capacity and structural integrity. Corrosion-induced expansion can also cause concrete to crack and spall, resulting in reduced serviceability [26]. In the case of RC bridges, corrosion significantly affects the rating factor and overall performance of the structure [13]. Furthermore, the study concludes that RC bridge decks experience increased longitudinal and transverse cracking due to overloads, which accelerates corrosion and structural deterioration [13]. A similar study on the service-life estimation of RC bridge structures exposed to chloride-induced reinforcement corrosion and variable traffic loads have been conducted. The findings indicate that the structural lifetime is significantly affected, with reductions ranging from 30% to 35%, and up to 70% in cases of high-frequency cyclic loading [27]. Hence, it is vital in analyzing how the bridge performs over time in relation to factors such as corrosion of reinforcing bars and variable load effects.

Since chloride-induced corrosion is primarily a concern in marine environments with high chlorine content in seawater [28], and Ethiopia is not exposed to such conditions, this study focuses on carbonation-induced corrosion as a dominant corrosion mechanism. Carbonation-induced corrosion with an exposure class of moderate humidity, i_{corr} of 0.10 to 0.20 μA/cm^2 [29] has been used. The attach penetration and the reduced diameter of reinforcement bars by corrosion are to be estimated from Eqs. (2) and (3), respectively [30].

$$P_x(t) = 0.0116\, I_{corr}\,(t - t_o)\ ,\ t > t_o \quad (2)$$

$$\phi(t) = \phi_0 - \alpha P_x(t) \quad (3)$$

Where, $\phi\ (t)$ is residual diameter at time t (mm), ϕ_0 is the initial bar diameter (mm), a is equal to 2 (for carbonated concrete), $P_x(t)$ is the average decrease of bar radius at time t, in mm, t_o is the time of corrosion initiation (years), t is elapsed time (years) and I_{corr} is the corrosion rate (μA/cm^2).

Corrosion of reinforcing bars significantly affects their mechanical properties, leading to a reduction in yield strength and it is computed from Eq. (4) [31]:

$$f_y(t) = \left(1 - \alpha\frac{A_s - A_s(t)}{A_s}\right) f_y \quad (4)$$

Where $f_y(t)$ is the residual yield strength of steel reinforcement, $A_s(t)$ is the residual area of steel reinforcement, A_s is an initial area of steel reinforcement and f_y is an initial yield strength of steel reinforcement and α is a yielding strength uncertainty coefficient (with a mean value of 0.5 and coefficient of variation of 12 %) [31].

The objective of this research was to evaluate the safety of reinforced girder bridges under various loading conditions with the specific focus on overloaded

vehicles. This research employs collected data from weighing stations to provide a systematic approach in evaluating the impact of heavy trucks on bridges. The conclusion made in this research revealed that overloaded vehicles have been identified to reduce the rating factors of RC bridges in Ethiopia by 23 %. The study also estimated that if the current trend of overloading and corrosion of reinforcing bars continues without the necessary monitoring and intervention, it would reduce the load carrying capacity of the bridges to 43 % as compared to their capacity under the legal loads. This significant reduction highlights the vulnerability of structural degradation over time. Thus, examining the effects of overloaded trucks on bridges and future possible corrosion of reinforcing bars aimed at raising awareness on the need for rigorous regulation and proper preventive measures towards enhancing the public safety.

2. METHODOLOGY

2.1 Data Collection

Traffic data, including truck loads, was acquired from three SWS sites; Modjo, Semera, and Sululta, over five-year period (Jan. 2018 to Dec. 2022) and more than 60,000 heavy trucks data were collected. Location of selected SWS sites is shown in Figure 1. The sites were chosen because there is a lot of heavy vehicle movement and those routes are where the majority of the country's economic transactions take place.

Figure 1 Location of weighing station sites.

2.2 Data Cleaning

The data were organized, filtered from erroneous records and the quality of the data was checked using the methodology and criteria provided in [32], [33]. Among the collected data, 51,900 were considered for analysis using the following criteria:

- trucks with a front axle weight of less than 3 tons were excluded as the minimum front axle weighs 3 ton to 4 ton [34], [35]
- vehicles with GVW $\geq 1.1\sum A_i$ or GVW $\leq 0.9\sum A_i$ were cleared [32]; where A_i is the i^{th} axle weight and GVW is the gross vehicle weight

An algorithm on Python software was developed to compute the impact of loads caused by moving vehicles. The program was prepared to produce all possible combinations taking into account varied bridge spans, axle arrangements, and axle weights and calculate the maximum load effects induced by vehicles. This tool enabled efficient simulation of critical load scenarios for different span lengths and the results were validated with hand calculations. To account future impacts, the live load was extrapolated for the remaining design period of the bridges. For the purpose of this research, the following methods were used; i) data collection and bridge inventory ii) traffic data analysis and iii) load rating analysis.

2.3 Bridge Selection

Bridge selection for evaluation is based on specific criteria to ensure a representative evaluation. In the present study, the Alemgena district has been selected as a critical area as it accounts for 74.3% of the collected truck load data, totaling 38,562 records. This substantial dataset provided a robust foundation for assessing the impact of traffic loads on bridges. Nine RC bridges from five road segments were selected with the following selection criteria:

- *Traffic exposure*: preference was given to bridges located along routes with high truck traffic intensity.
- *Age of structure*: bridges over 50 years old were prioritized to evaluate deterioration patterns and the impact of aging on structural load capacity. In addition, relatively newer bridges with an age of around 10 years were also included to enable performance comparison across different service life stages.
- *Span length*: bridges with a span length of 15 meters (each span) and above were selected, as this length accommodates the full front-to-rear axle spacing of a 7-axle truck, which measures approximately 15.0 m as indicated in Table 1.
- *Span type*: different span types were considered.
- *Structural type consistency*: only RC girder bridges were selected to ensure uniformity in load analysis and rating methodology.

The identified bridges and some of the pictures from ERA Bridge Management System (BMS) [36] are shown in Table 1 and Figure 2, respectively.

Table 1 Bridge data

Bridge Id.	Bridge name	Bridge length (m)	Span support type	Span composition	Construction year
A1-1-004	Gogecha	57.9	Multiple	3×19.3	2014
A1-2-008	Unnamed	42	Multiple	2×21	2013
A1-3-006	Mermersa	18	Single	18	1974
A3-1-024	Duber Guda	41	Multiple	2×20.5	1967
A4-2-025	Quribe	24	Single	1×24	1982
A7-1-001	Koka	55.55	Multiple	3×18.5	1953
A7-1-002	Awash	93.5	Multiple	5×18.7	1953
A7-1-003	Meki	20	Single	1×20	1961
A5-1-013	Awash	52	Continuous	9+34+9	1980

(a) (b)

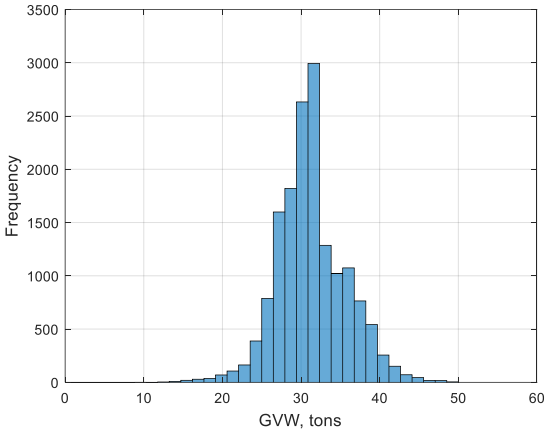
Figure 2 RC bridges a) Gogecha bridge and b) Koka bridge.

3. RESULTS AND DISCUSSION

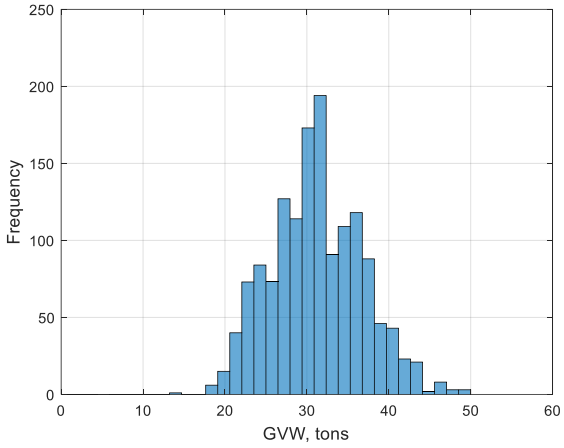
3.1 Statistical Properties of SWS Data

The Ethiopian Standards Agency (ESA) calibrated the weight measuring devices, and the quality control processes ensured the data was complete, axle load measurements were accurate, and the results were validated against manual inspection and calibration standards.

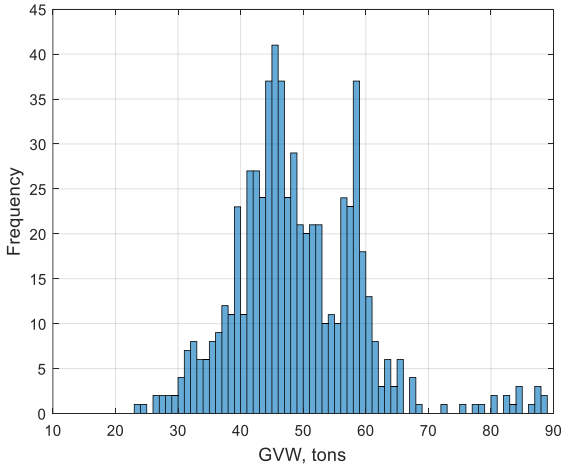
Hence, in this study, no additional checks on sensor stability and data integrity were made. The data were grouped and classified as per axle, and then a field survey was undertaken for the determination of axle configurations. The histogram plot of Gross Vehicle Weight (GVW) is shown in Figure 3.



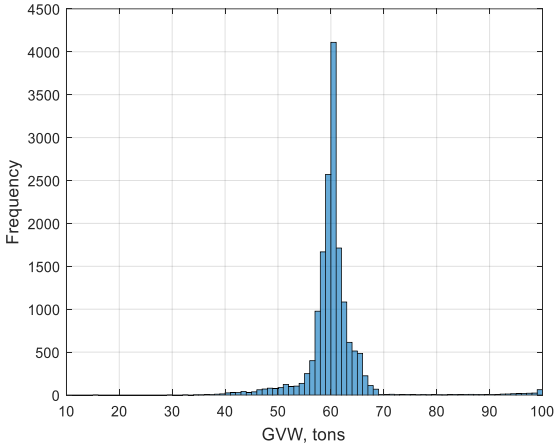
(a)



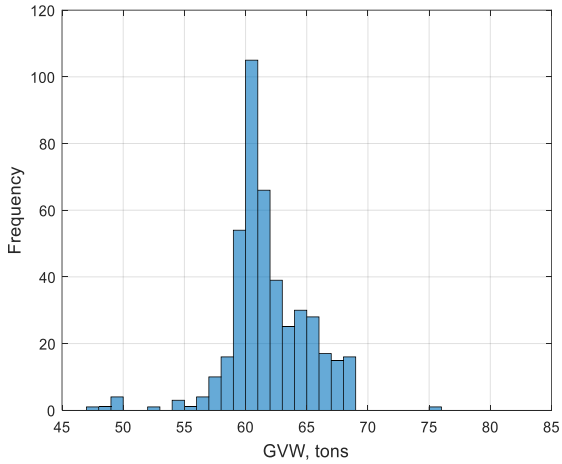
(b)



(c)



(d)



(e)

Figure 3: Histogram plot of GVW for (a) 3-axles (b) 4-axles (c) 5-axles (d) 6-axles and (e) 7-axles.

A diagram of a 7-axle vehicle configuration with a front axle designated as A1 is shown in Figure 4 and field measurements of axle spacings are shown in Table 2.

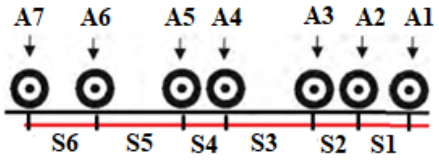


Figure 4: Schematic diagram of vehicle and axle load configurations.

3.2 Bridge Load Limit Requirements

3.2.1 National Law Requirement

Axle loads and GVW that violate load restrictions set forth in Ethiopian

regulations and other standards are examined and the results are shown in Table 3. As shown in Table 3, for the case of 3-axle vehicles, among the 18,070 vehicles data, 9,938 vehicles (55%) exceed the axle load limit specified in the regulation of the country [37]. However, when it comes to GVW, it's worth noting that 44.9% of the 5-axles trucks exceed the national law's requirements [37]. The data shows a truck with 7-axles weighs a gross weight of 124-ton, which is a heavily loaded truck along the selected routes. A study on vehicle data collected from weigh stations across the East and Southern African regions (ESA) was conducted, and the results showed that the issue of vehicle overloading and the urgent need for effective control measures have long been recognized. However, due to various challenges, efforts to manage overloading have largely remained ineffective. According to the report, the incidence of overloaded trucks in the ESA region ranges from 10% to 50% [38], which aligns with the findings of the present study, where it ranged from 3.2% to 55%, as shown in Table 3.

Table 2 Axle spacing measurement

No. of axles	Values	Axle spacing (m)					
		S1	S2	S3	S4	S5	S6
3	Max.	4.3	1.3	–	–	–	–
	Min.	3.2	1.25	–	–	–	–
4	Max.	4.74	4.58	4.6	–	–	–
	Min.	1.7	2.3	1.4	–	–	–
5	Max.	3.35	4.4	6.9	1.6	–	–
	Min.	3.16	1.3	6.5	1.4	–	–
6	Max.	4.2	1.4	6.14	3.8	1.4	–
	Min.	3.36	1.34	4.8	1.34	1.35	–
7	Max.	4.2	1.36	5.42	1.4	1.35	1.32
	Min.	3.23	1.25	4.8	1.38	1.22	1.28

.

Table 3 Number of illegal vehicles and bridge load limits violations.

No. of axles	No. of vehicles	Overloaded or illegal vehicles (National law, Negarit)					
		Axle load (ton)			GVW (ton)		
		Axle load limit	No.	%	GVW, limit	No.	%
3	18,070	8 - 10	9,938	55.0	28	5,819	32.2
4	1,915	8 - 10	411	21.5	38	120	6.3
5	3,540	8 - 10	893	25.2	48	1,589	44.9
6	27,720	8 - 10	7,762	28.0	58	892	3.2
7	655	8 - 10	165	25.2	68	42	6.4
	51,900		**19,169**	**36.9**		**8,462**	**16.3**

3.2.2 Regional and International Standard Requirements

In addition, the requirements for Federal Bridge Formula B (BFB) [39], [40] and the Tripartite Transport and Transit Facilitation Program (TTTFP) [41] to limit the permissible weight that a bridge can sustain were checked. For checking, Eqn. (5) [39], [40] and Eqn. (6) [41] have been used. The number of truck data exceeding the BFB and TTTFP limits are presented in Table 4.

$$W = 0.75\left(\frac{LN}{N-1} + 3.65N + 11\right) \quad (5)$$

Where W is maximum weight that can be carried on a group of two or more axles (ton), L is the distance between the outer axles of any two or more consecutive axles (m) and N is number of axles being considered.

$$W = 2100L + 18000 \quad (6)$$

Where W is the permissible mass (kg) and L is the distance between two axles (m).

From Table 4, it is observed that, among the total trucks considered, 40.2 % and 27.49 % of the BFB and TTTFP bridge formulas are violated, respectively. Hence, without controlling the movement of these types of trucks, the bridges will accumulate more damages and incur additional cost for maintenance intervention. Therefore, uncontrolled growth in loads and volumes of heavy trucks should be monitored, reassess bridges' strength accounting the current traffic conditions that actually exist in Ethiopia is found to be necessary.

In Tables 3 and 4, it is shown that more than 16% of the trucks were loaded beyond the weight limits set by the national regulations, and bridge formulas. This can compromise the efficiency and stability of the structure over time. Thus, addressing these issues on a regular basis and implementing regulatory enforcement are important to ensure safety of bridges.

Table 4 Trucks violating the bridge formulas

No. of axles	No. of vehicles	Trucks violating the bridge formulas					
		BFB			TTTFP		
		GVW, limit (ton)	No.	%	GVW, limit (ton)	No.	%
3	18,070	22.3	8,931	49.42	28.9	5,272	29.18
4	1,915	26.8	682	35.61	34.1	403	21.04
5	3,540	34.9	961	27.15	47.1	714	20.17
6	27,720	38.2	10,043	36.23	49.8	7,661	27.64
7	655	40.1	247	37.71	48.6	215	32.82
	51,900		**20,864**	**40.20**		**14,265**	**27.49**

3.3 Strength Evaluation

Utilizing load rating methodologies specified by ERA and AASHTO bridge evaluation manuals and the structural capacity of bridges under various loading conditions was found important. For bridge rating, Eq. (7) was used [1], [5].

$$RF = \frac{\varphi R_n - \gamma_{Di} D_i - \gamma_{DW} DW}{\gamma_{Li}(L_i + I)} \quad (7)$$

Where RF is the rating factor, φR_n is the nominal resistance $=A_s f_y(d-a/2)$, D_i is the effect of dead loads, L_i is the live-load effect for load i other than the rating vehicle, L_i is the nominal live-load effect of the rating vehicle, I is the impact factor for the live-load effect, γ_{Di} is the dead load factor, γ_{Li} is the live load factor.

Even though the principles of safety evaluation of bridges are generally the same, the requirements and norms governing the design and assessment of bridges may be different. These may include load factors, material specification, design codes, and safety margins which may be affected by the geographical area, historical practice, and improvement in engineering standards. Bridges constructed in different years have different remaining service lives; the extent of the deterioration of the material performance and the degradation of the overall structural performance are also different [1]. As a result, the corresponding material strengths and deterioration rates are considered as a criteria for the safety evaluation of bridges.

3.3.1 Bridge Data

The dimensions of the bridges used in this study were taken from ERA BMS and they were used to compute the effect of dead load of the bridge. Table 5 shows the bridge data used in the study [36].

Table 5 Bridge dimensions and damage

Bridge Id.	Span (m)	Slab thickness (cm)	No. of girders	Girder depth (m)	Girder spacing (m)	Girder width (m)	Damage %
A1-1-004	3×19.3	20	4	1.3	2.2	0.4	2.76
A1-2-008	2×21	25	5	1.35	3.1	0.4	1.88
A1-3-006	18	25	5	1.35	3.1	0.4	9.31
A3-1-024	2×20.5	20	4	1.35	2.1	0.4	11.69
A4-2-025	1×24	20	7	1.1	1.04	0.3	12.31
A7-1-001	3×18.5	20	4	1.0	1.75	0.4	14.2
A7-1-002	5×18.7	20	4	1.2	1.8	0.4	10.3
A7-1-003	1×20	20	4	1.3	2	0.45	10.67
A5-1-013	9+34+9	50	4	2.4	2.5	0.4	0.89

3.3.2 Loading Conditions

For strength evaluation of bridges, manuals specify the legal truck load in terms of number of axles, axle configuration, and axle load [1], [5]. For the computation of effect of live load, these truckloads were used. Furthermore, the live load data obtained from SWS were considered.

3.3.3 Effect ofL

The concept of influence lines was used to compute effects of loads. This approach enabled to evaluate corresponding maximum load effects in terms of shear force and bending moment. The analysis focused on bridges with single lanes loaded, and for longer spans, the impact of multiple vehicles with a specified headway distance between successive trucks was considered. Since the collected data were from SWS, headway distances were measured on-site, and various combinations of truck arrangement with headway distances were considered accordingly. The statistical distribution of headway distances showed a mean distance of 7.51 m and a standard deviation of 2.59 m, with values ranging from 3.28 m to 12.01 m.

For the analysis of the loading effects of the vehicles for a specific axle arrangement, a computer program was developed to consider all possible combinations (collected from the weighing stations). The effects of dead and live loads are shown in Table 6.

Table 6 Effects of dead and live loads

Bridge Id.	Dead load effects		Legal load effects			Effects of actual truck loads	
	Shear (kN)	Moment (kN-m)	Shear (kN)	Moment (kN-m)	Governing Load	Shear (kN)	Moment (kN-m)
A1-1-004	251.58	1,213.85	270.00	941.62	Legal load 1	609.31	2,499.19
A1-2-008	379.42	1,991.94	283.00	1,089.70	Legal load 2	642.31	2,867.77
A1-3-006	325.22	1,463.47	259.41	837.49	Legal load 2	583.94	2,232.98
A3-1-024	264.91	1,357.67	280.30	1,037.50	Legal load 2	632.62	2,756.96
A4-2-025	139.82	838.94	302.31	1,406.70	Legal load 3	700.14	3,574.82
A7-1-001	185.16	856.37	264.05	877.51	Legal load 1	597.33	2,333.77
A7-1-002	205.12	958.92	265.83	893.53	Legal load 2	597.60	2,374.65
A7-1-003	259.8	1,299.00	276.57	997.77	Legal load 3	622.91	2,648.15
A5-1-013	853.51	6,448.75	255.12	1,848.15	Legal load 3	740.87	4,964.26

3.3.4 Section Capacity of Bridges

In the absence of specific details on reinforcement and material grades for certain bridges, estimations of reinforcing bars (longitudinal and transversal) are made based on the assumption that the bridges were designed following standard manuals and they comply with the load requirement with appropriate material specifications defined in bridge design manuals [5], [6]. Furthermore, it was assumed that the construction was carried out according to the design specifications, using the specified materials, with proper detailing, and done by a qualified contractor [42].

In load rating calculations, variations in bridge age and the historical codes under which each bridge was originally designed presented a significant challenge. To address this, the study employed the following approaches:

- where original material specifications differed from current standards, material grades using conservative estimates were used as specified in [1], [5], [42].
- for bridges without plans, the area of reinforcing bars was estimated as a percentage of the gross area of the beams, provided there were no indications of significant distress [42]. Alternatively, assumption of original design were done by back-analysis or redesign [43].
- despite variations in original design procedures, the AASHTO LRFR method could be applied to uniformly rate all bridges [42].
- damage conditions of the bridge were to be used [44]. In the present study, damage rates obtained from the bridge data base [36] were considered.

Regarding materials property, the compressive strength of concrete and steel yield stresses used for various types of reinforcing steel grades are given in Table 7. In most cases, the stress of the concrete was assumed to be 20.7 MPa and that of the steel as f_y=276 MPa. For newly constructed bridges, f_y of 314 MPa has been used. For deteriorated bridges, f_y=228 MPa, 248 MPa and for concrete a strength of 15 MPa were considered [1].

Using the above assumptions, the bridges were redesigned and their section capacities were evaluated using Response 2000 software, which uses the modified compression field theory [45]. The estimated reinforcing areas, material properties and section capacities of the bridges are presented in Table 7.

Table 7 Input data and section capacity of bridges

Bridge Id.	A_S (mm^2)	Stirrups (mm)	Yield stress, f_y (MPa)	f'_c (MPa)	Estimated section capacity	
					Shear (kN)	Bending (kN-m)
A1-1-004	8,844	ϕ 12 c/c 150	314	20.7	1,139.8	3,717
A1-2-008	12,060	ϕ 12 c/c 130	314	20.7	1,331.6	4,536
A1-3-006	12,060	ϕ 12 c/c 120	276	20.7	1,168.9	4,329
A3-1-024	10,452	ϕ 12 c/c 150	248	15.0	977.6	3,119
A4-2-025	19,296	ϕ12 c/c 170	276	20.7	830.96	3,137
A7-1-001	10,452	ϕ 12 c/c 230	228	20.7	626.2	2,034
A7-1-002	11,256	ϕ 12 c/c 210	228	20.7	740.24	2,817
A7-1-003	11,256	ϕ 12 c/c 210	248	20.7	791.98	3,310
A5-1-013	17,190	ϕ 12 c/c 110	276	20.7	1,954.6	10,950

The reinforcement bars were assumed to have a diameter of 32 mm. However, no ground-penetrating radar or cover-meter measurements were conducted to validate

this assumption, and the influence of bar diameter on section capacity was not included. Sample material properties and cross section details of Gogecha Bridge (A1-1-004) are shown in Figure 5.

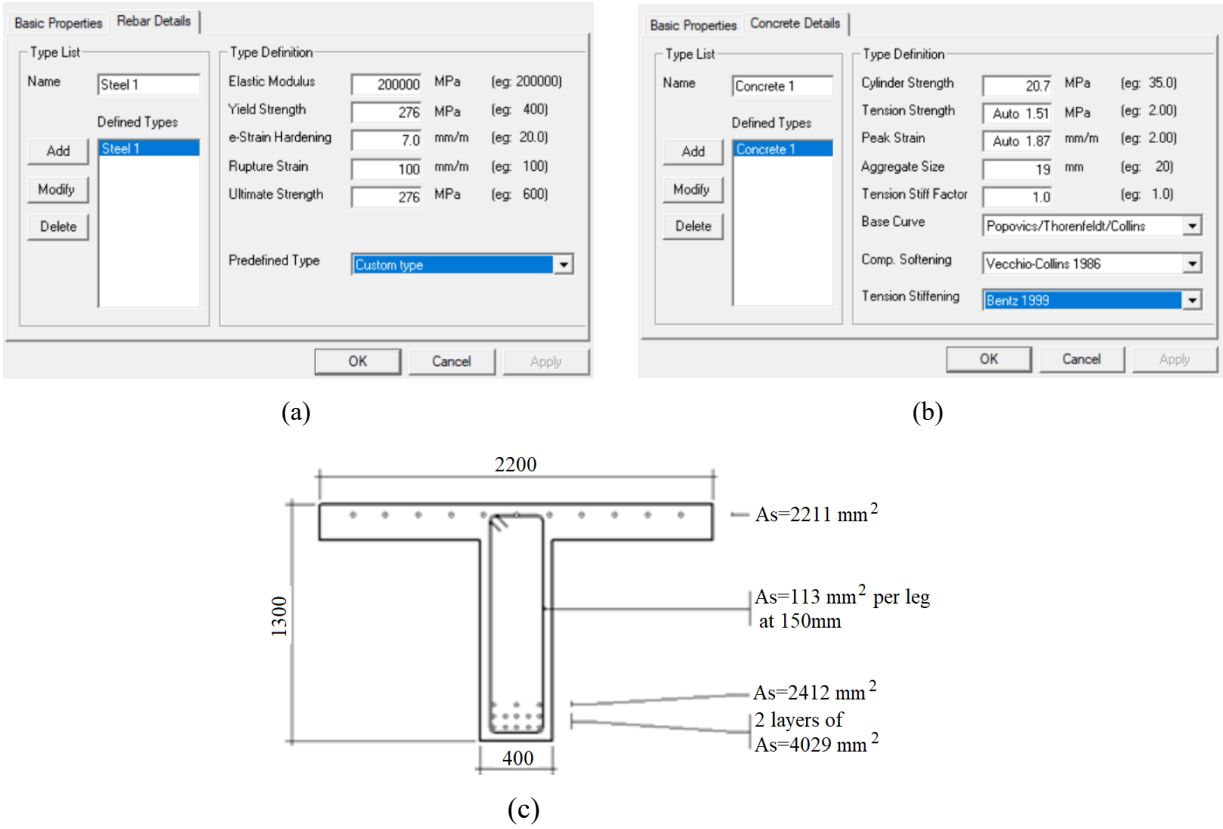


Figure 5: (a) rebar properties (b) concrete properties (c) cross section details

3.3.5 Factors for strength evaluation

In the computation of rating factors for legal loads, the load factors were set at γ_D= 1.2 and γ_L = 1.65. However, for vehicles that exceed legal loading limits, these factors were reduced (in this study, γ_D and γ_L values of 1.05 were used), reflecting the need for flexibility in evaluating the structural capacity under more extreme conditions [46]. As per the recommendation of the manuals [1], [5], the resistance factors were set at 0.80 for deteriorated bridges and 0.95 for those in good condition. These values were vital in accounting for the current state of the bridge's materials and construction quality. Since all the bridges considered in this study were of two lanes, reduction factors for live load of 1.0 has been used. In addition, for all bridges, an impact factor of 0.1 (fair condition of wearing surface was reported [36]) and condition factor of 1.0 have been used [5]. Field test results showed that the dynamic amplification factor (DAF) for bridges depends on loaded length, vehicle speed and pavement condition [47]. However, in this study, a constant value of impact factor was used for all bridges. The live load distribution factors specified in the AASHTO LRFD (Load and Resistance Factor Design) specifications were used for strength evaluation [1], [14] and are generally conservative in most cases, especially for straight bridges [14]. The combination of load factors, resistance factors, impact factors, and distribution factors provided a comprehensive framework for evaluating bridge strength. This holistic approach was important for ensuring that all aspects of bridge performance were considered. The distribution factors (DF) are given in Table 8.

3.3.6 Bridge ratings

3.3.6.1 Current condition

The rating factors of bridges due to legal and actual truck load data were computed using Eqn. (7) and the results are summarized in Table 8.

.

Table 8 Rating factors for shear and moment

Bridge Id.	DF		Rating factors for legal trucks			Rating factors for actual loads			
	SF	BM	SF	BM	RF_{Legal}	RF_{SF}	RF_{BM}	RF_{Actual}	RF hierarchy
A1-1-004	0.80	0.62	1.99	1.96	1.96	1.45	1.26	1.26	2
A1-2-008	0.80	0.79	1.97	1.23	1.23	1.46	0.85	0.85	6
A1-3-006	0.80	0.81	1.76	1.74	1.74	1.32	1.13	1.13	3
A3-1-024	0.77	0.64	1.43	0.98	0.98	1.07	0.68	0.68	8
A4-2-025	0.49	0.36	2.16	1.98	1.98	1.52	1.31	1.31	1
A7-1-001	0.68	0.54	1.05	0.93	0.93	0.79	0.64	0.64	9
A7-1-002	0.68	0.54	1.28	1.58	1.28	0.96	1.03	0.96	5
A7-1-003	0.75	0.70	1.07	1.13	1.07	0.82	0.76	0.76	7
A5-1-013	0.89	0.62	2.02	1.28	1.28	1.26	1.02	1.02	4

As shown in Table 8, under legal load conditions, the rating factors for two bridges (A3-1-024 and A7-1-001) are found to be less than 1.0, indicating there is a need to establish the LRFR posting load restriction [1], [5], [48]. However, these values do not indicate severe structural failure [49]. The deviations of these bridges from the standard are concerns for their structural capacity and the potential need for further investigation or reinforcement to ensure safety [1]. Table 8 presents the rating factor hierarchy, arranged from the highest RF to the lowest. This hierarchy serves as a systematic tool for identifying bridges with lower rating factors as priority cases, thereby supporting decisions on maintenance, rehabilitation, or replacement.

Actual truck data gathered from weighing stations, in all cases, reduces bridges' performance (rating factors) as compared to legal loads and shortens their service life. As a result of the overloaded trucks, the rating factors of seven bridges were now found to be less than 1.0. The identification of such reduced rating factors emphasizes the importance of continuous monitoring and assessment of bridge conditions to maintain infrastructure safety.

3.3.6.2 Future condition of bridges

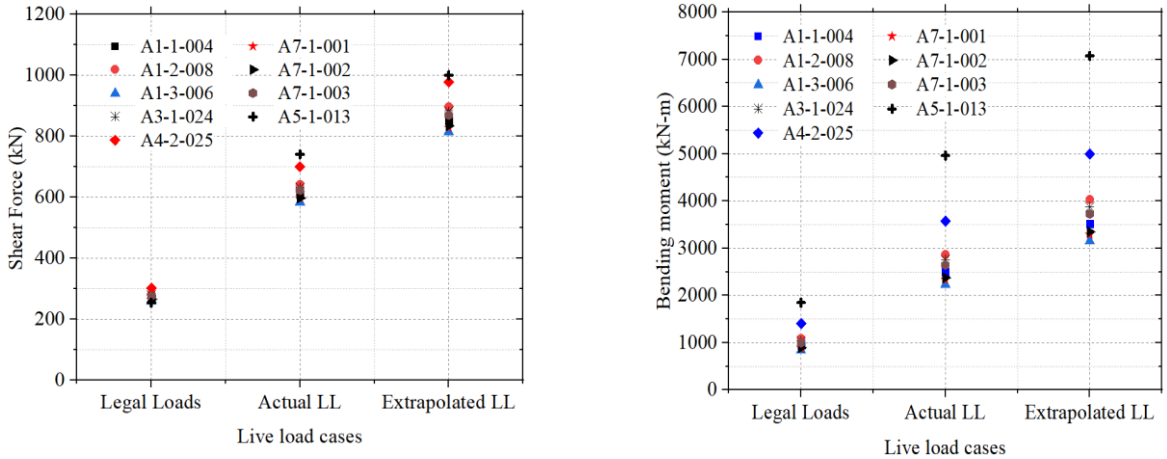
To predict bridge's performance in the long-run, factors like deterioration rates and extrapolated loads which the bridges are likely to experience in the future need to be considered [7, 26]. For the calculation of extrapolated load data for a return period of 75 years, Eqn. (1) is used. In line with this, the extrapolated bending moment and shear force for the remaining service period are computed and shown in Table 9. Estimating the remaining service period of a bridge involves considering its original design life and the year it was constructed. This estimation is a fundamental step in bridge management and provides a basis for further assessments and informs maintenance interventions. Considering corrosion of reinforcing bars and reduced yield strength of rebars, the reduced section capacity of bridge was recalculated accordingly. Consequently, bridge rating factors for both shear and moment have been computed based on the updated values of extrapolated live loads and reduced section capacity as shown in Table 9. In this study, to account future deterioration, the resistance factors were reduced by 0.05. In Figure 6, comparison of live load effects is shown. The extrapolated live load effects demonstrated that a vehicle causes an

increment of 40% on average for shear force and bending moment as compared to the current traffic with a load increment rate of 0.005 per year was obtained, showing that the live load intensity will increase by 40% over a 75-year period, which is in the same range with research carried by Wang and Li [50]. The analysis result of the current study also showed significant increments of live load effects, which were 56.9 % for shear force and 62.2 % for bending moment as compared to the effect of legal loads.

Table 9 Effects of corrosion rate on section capacity and extrapolated live loads

Bridge Id.	$T_{rem,ser}$*	Reduced rebar area (mm^2)	Reduced yield stress f_{yR} (MPa)	Reduced section capacity		Extrapolated live load effects		RF_{SF}	RF_{BM}	$RF_{Extrap.}$
				SF (kN)	BM (kN-m)	SF (kN)	BM (kN-m)			
A1-1-004	65	8,676.35	306.86	1,095.83	3,573.61	837.21	3,508.88	0.93	0.77	0.77
A1-2-008	64	11,834.85	306.97	1,280.61	4,362.29	880.62	4,015.26	0.93	0.50	0.50
A1-3-006	25	11,970.34	273.54	1,137.01	4,210.88	710.12	3,046.44	0.95	0.72	0.72
A3-1-024	18	10,395.44	246.39	952.86	3,040.08	742.00	3,737.60	0.81	0.42	0.42
A4-2-025	33	19,107.98	272.77	806.41	3,044.30	888.30	4,907.49	1.07	0.84	0.84
A7-1-001	4	10,437.80	227.63	612.84	1,990.61	776.53	3,119.83	0.54	0.41	0.41
A7-1-002	4	10,437.80	227.63	724.45	2,756.91	718.32	3,147.16	0.71	0.68	0.68
A7-1-003	12	11,214.63	246.91	773.29	3,232.08	747.49	3,543.79	0.60	0.49	0.49
A5-1-013	31	17,032.42	272.96	1,897.95	10,632.63	900.46	6,972.83	0.88	0.56	0.56

*$T_{rem,ser}$ = remaining service period



(a) (b)

Figure 6 Comparison of live load effects (a) shear force and (b) bending moment

3.3.6.3 Summary of rating factors

In Table 10, summary of bridge ratings for legal loads, actual truck data and extrapolated live loads are shown. In the table, rating factors for legal loads considering deterioration of the bridge due to corrosion are indicated. It is also noted that most of the existing RC girder bridges do not meet the current standards for modern traffic loads. A consistent downward trend is observed from legal to actual to extrapolated rating factors across all bridge cases. The percentage reduction between RF_{Legal} and RF_{Actual} averages 30.18 %, indicating the impact of current loading conditions on structural performance.

On the other hand, the extrapolated live loads give rating factors ($RF_{Extrap.}$) below one for all bridges. The results indicate a significant reduction of 56.29 %, even in the absence of material deterioration of

concrete and other environmental conditions which lead the bridges to fail. This result is relatively higher than that of a related study, which reported a 36.5 % reduction in reliability index under overloaded traffic conditions [51]. It was also observed that when steel bars are corroded in the future and legal load effects are considered, the rating factors ($RF_{Legal,corr}$) are, on average, 17.71 % greater than those based on current truck load data without corrosion (RF_{Actual}). This shows that overloading causes a more immediate reduction in section capacity while corrosion progressively deteriorates the steel reinforcement, resulting in a gradual reduction in section capacity and becoming a critical concern over time.

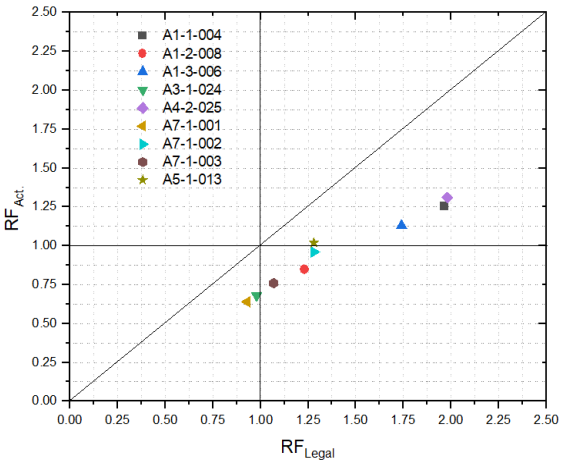
A comparison of rating factors under legal loads for both current and future condition is shown in Table 10. The result shows that, when considering the effect of corrosion alone, anticipated bridge conditions have an average reduction of 18.0 % in rating factors compared to their current performance. This falls within the findings of El Maaddawy et al. [52], where corrosion led to a reduction in section capacity ranging from 6.5 % to 29 % under sustained load.

Table 10 Summary of rating factors

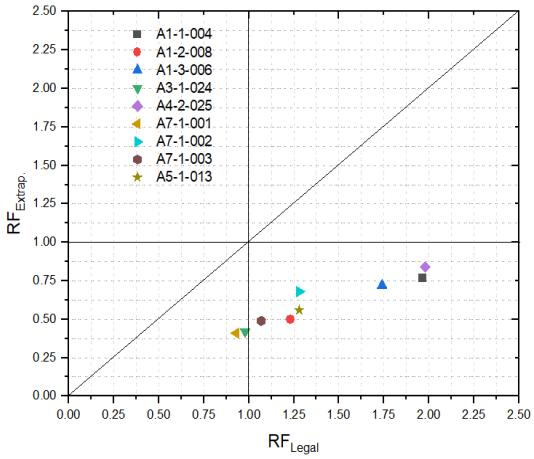
Bridge Id.	RF_{Legal} (1)	RF_{Actual} (2)	$RF_{Extrap.}$ (3)	$RF_{Legal,corr}$ (4)	% reduction (1) and (2)	% reduction (1) and (3)	% reduction (1) and (4)	% reduction (2) and (4)
A1-1-004	1.96	1.26	0.77	1.66	35.71	60.71	18.07	24.10
A1-2-008	1.23	0.85	0.50	0.98	30.89	59.35	25.51	13.27
A1-3-006	1.74	1.13	0.72	1.48	35.06	58.62	17.57	23.65
A3-1-024	0.98	0.68	0.42	0.79	30.61	57.14	24.05	13.92
A4-2-025	1.98	1.31	0.84	1.72	33.84	57.58	15.12	23.84
A7-1-001	0.93	0.64	0.41	0.77	31.18	55.91	20.78	16.88
A7-1-002	1.28	0.96	0.68	1.13	25.00	46.88	13.27	15.04
A7-1-003	1.07	0.76	0.49	0.92	28.97	54.21	16.30	17.39
A5-1-013	1.28	1.02	0.56	1.15	20.31	56.25	11.30	11.30
Average					**30.18**	**56.29**	**18.00**	**17.71**

The comparative plot of the rating factors for bridge is shown in Figure 7. Most data points lie below the reference line, indicating that the majority of bridges experience a reduction in rating factors when subjected to current and extrapolated traffic load effects. Results show a consistent decline from legal ratings to actual and extrapolated values, indicating progressive structural degradation over time, raising concerns about their structural safety. Similarly, Figure 8 presents a comparison of bridge ratings under various load cases.

The trends of rating factors of the present study show a significant decrease in load-carrying capacity over time, emphasizing the need for proactive maintenance, monitoring, and potential strengthening interventions. Variations in bridge characteristics and exposure conditions are required for bridge-specific assessments to ensure accurate management planning.



(a)



(b)

Figure 7: Comparison of rating factors a) RF_{Legal} vs. RF_{Actual} and b) RF_{Legal} vs. $RF_{Extrap.}$

In summary, the findings of this study highlighted the significance of controlling excessive truck loads to ensure the safety and capacity performance of highway bridges. In situations where the percentage of overloaded vehicle is significant, and to reflect their actual effect on bridge structures, researchers recommend calibration of live load models and load factors.

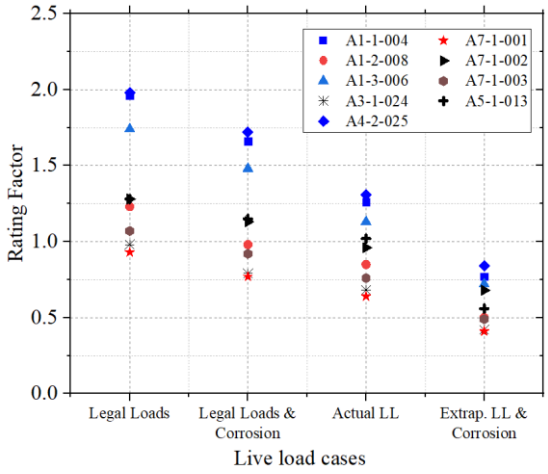


Figure 8 Bridge ratings for different cases.

Accordingly, efforts have been made toward developing design live load models [7], [17], [21], [32], [53] and calibrating load and resistance factors for bridge design and evaluation [7], [54], [55], [56]. This calibration aimed to reflect realistic loading scenarios more accurately, considering the variations and trends observed in current traffic patterns, particularly due to vehicle overloading and aging infrastructure.

4. CONCLUSIONS

A study on the effect of overloaded trucks on selected RC highway bridges in Ethiopia revealed that current traffic loading exceeds legal limits by 56.9% in shear force and 62.2% in bending moment, with vehicle overload percentages of 16.3%, 40.2%, and 27.49% compared to national, BFB, and TTTFP regulations, respectively. The absence of traffic monitoring has led to severe bridge damage, reducing rating factors by 30.18% and accelerating deterioration. The findings highlight the need for cost-benefit analysis to assess projected economic impacts, probabilistic assessments to account uncertainties of random variables (traffic growth, material degradation, climate change), and calibration of live load models, load and resistance factors. While extrapolated load analyses are valuable tools for bridge safety evaluations, their predictions should be supported by field monitoring, probabilistic load models, and sensitivity analyses.

Effective monitoring of overloaded truck is often lacking in the country. This issue can only be remedied through collaboration between regulatory bodies, law enforcement, and bridge owners. Therefore, it is important for transportation agencies to enforce regulatory standards and polices aimed at reducing impacts of trucks with excessive loads on bridge performance and safety.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest with others.

ACKNOWLEDGEMENTS

The authors thank the Ethiopian Roads Administration (ERA) for providing access to the axle weight data.

REFERENCES

[1] AASHTO 2018, "Manual for Bridge Evaluation," *American Association of State Highway and Transportation Officials*, 2018.

[2] J. Zhou, X. Ruan, X. Shi., and C. C. Caprani, "An Efficient Approach for Traffic Load Modelling of Long Span Bridges," *Structure and Infrastructure Engineering*, vol. 15, no. 5, 2019, pp. 569-581.

[3] K. Dey, and M. Braja, "Risk-Based Bridge Management Systems: Overview, Application, and Future Trends," *Journal of Bridge Engineering*, vol. 24, no. 1, 2019.

[4] N.J. Bertola and E. Brühwiler, "Risk-based Methodology to Assess Bridge Condition based on Visual Inspection," *Structure & Infrastructure Engineering*, vol. 19, no. 4, 2023, pp.575-588.

[5] ERA-2013, "Bridge Design Manual," *Ethiopian Roads Authority*, Addis Ababa, Ethiopia, 2013.

[6] AASHTO 2017, "LRFD Bridge Design Specifications, American Association of State Highway and Transportation Officials, " 4th ed., 2017.

[7] B. Wesenseged, "Development of Live Load Model for Ethiopian Short to Medium Span Bridges," MSc. Thesis, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia, 2021.

[8] S. Gabriel, G.L. Jaime, W. Chao and P. Mert, "Condition Assessment of Existing Bridges: A Case Study of The Kalix Bridge," *Research Report,* Luleå University of Technology, 2024.

[9] T. Omar and M.L. Nehdi, "Condition Assessment of Reinforced Concrete Bridges: Current Practice and Research Challenges," *Infrastructures*, vol. 3, no. 3, 2018.

[10] S. Lee, N. Kalos and D. H. Shin, "Non-destructive Testing Methods in the US for Bridge Inspection and Maintenance," *KSCE Journal of Civil Engineering*, vol. 18, 2014, pp.1322-1331.

[11] A. O'Connor, and E. M. Eichinger, "Site-Specific Traffic Load Modelling for Bridge Assessment," *Proceedings of the Institution of Civil Engineers - Bridge Engineering*, pp. 185-194, 2007.

[12] R. W. James, R. A. Zimmerman, and C. R. McCreary, "Effects of Overloads on Deterioration of Concrete Bridges," *Transportation Research Record*, vol. 118, 1987, pp. 65-72.

[13] S. U. Khan, T. Ayub and A. Qadir, "Effect of Overloaded Vehicles on the Performance of Highway Bridge Girder: A Case Study," *Procedia Engineering*, vol. 77, pp. 95–105, 2014, doi: 10.1016/j.proeng.2014.07.010.

[14] S. L. Wood, M. J. Hagenberger, B. E. Heller, and P. J. Wagener, "Evaluation of Serviceability Requirements for Load Rating Prestressed Concrete Bridges," *Center for Transportation Research*, Report No. FHWA/TX-07/0-1895-1, 2007.

[15] B. Sivakumar and E. Senturk, "Bridge Load and Resistance Factor Rating (LRFR) Assessment - Statewide," New York Engineering and Architecture, P.C., Report No. C-11-04, 2020.

[16] E.J. O'Brien, B. Enright and A. Getachew, "Importance of the Tail in Truck Weight Modeling for Bridge Assessment," *Journal of Bridge Engineering*, vol. 15, no. 2, 2010, pp. 210-213.

[17] P. van der Spuy and R. Lenner, "Towards a New Bridge Live Load Model for South Africa," *Structural Engineering International*, vol. 29, no. 2, 2019, pp. 292-298.

[18] B. Enright, "Simulation of Traffic Loading on Highway Bridges," Ph. D Dissertation, College of Engineering, Mathematical and Physical Sciences, University College Dublin; 2010.

[19] A. Ferreira and L. De Haan, "On the Block Maxima Method in Extreme Value Theory: PWM Estimators," *The Annals of Statistics*, 2015, pp.276-298.

[20] C.C. Caprani, "Probabilistic Analysis of Highway Bridge Traffic Loading," PhD Dissertation, School of Architecture, Landscape and Civil Engineering, University College, Dublin, 2005.

[21] A.S. Nowak, and P. Rakocz, "WIM-based Live Load for Bridges,", KSCE *Journal of Civil Engineering*, vol. 17, 2013, pp. 568-574.

[22] B. Sivakumar, M. Ghosn, and F. Moses, "Protocols for Collecting and using Traffic Data in Bridge Design," *Transportation Research Board*, vol. 683, 2011.

http://onlinepubs.trb.org/onlinepubs/archive/notesdocs/nchrp12-76_fr.pdf

[23] X. Zhou, "Statistical Analysis of Traffic Loads and Their Effects on Bridges, using Weigh-In-Motion Data Collected in France," PhD Dissertation, Université Paris-Est, 2013.

[24] X. Zhou, F. Schmidt, and B. Jacob, "Extrapolation of Traffic Data for Development of Traffic Load Models: Assessment of Methods used during Background Works of Eurocode," 6th *Conference on Bridge Maintenance, Safety and Management*, 2012.

[25] B. Sivakumar, M. Ghosn, and F. Moses, "Adjustment of Load and Resistance Factor Design Live Load Factors using Recent Weigh-in-Motion Data," *Transportation Research Record*, vol. 2200, no. 1, 2010, pp. 90-97.

https://doi.org/10.3141/2200-11.

[26] A.G. Tarekegn, "Time-Dependent Reliability Analysis for Deflection of a Reinforced Concrete Box Girder Bridge" *International Journal of Engineering and Applied Sciences*, vol. 15, no. 2, 2023, pp. 60-74.

[27] M.M. Borah and A. Sil, "Service-Life Estimation of a Reinforced Concrete Bridge Structure Exposed to Chloride-Contaminated Environments and Variable Traffic Loads," ASCE-ASME *Journal of Risk and Uncertainty in Engineering Systems*, Part A: Civil Engineering, vol. 9, no. 4, 2023.

[28] Y. Tian, Y. Liu, Y. Xie, G. Y. Zhang, H. Ye, D. Yan, and Q. Cai, "Coupled Effects of Chlorides and Sulfates on Steel Reinforcement Corrosion in Concrete Structures: A Comprehensive Review," *Case Studies in Construction Materials*, vol. 22, 2025.

[29] C. Wang, "Structural Reliability and Time Dependent Reliability," *Reliability Engineering*, Springer, 2021.

[30] D. Baingo, "A Framework for Stochastic Finite Element Analysis of Reinforced Concrete Beams affected by Reinforcement Corrosion," PhD Dissertation, University of Ottawa, Canada, 2012.

[31] H. Travis, "Reliability of Corroding Reinforced Concrete Structures," M.Sc. Thesis, B.S., Louisiana State University, 2019.

[32] A. L. Carneiro, E. D. L. Portela and T. N. Bittencourt, "Development of

Brazilian Highway Live Load Model for Unlimited Fatigue Life," *Revista IBRACON de Estruturas e Materials*, vol. 13, no. 4, 2020.

[33] T. Wei and J. Fricker, "Weigh-In-Motion Data Checking and Imputation," FHWA/IN/JTRP-2003/16, 2470; 2003.

[34] O. Selezneva and D. Wolf, "Successful Practices in Weigh-in-Motion Data Quality with WIM Guidebook," Arizona Department of Transportation, vol. 1 & 2, FHWA-AZ-17-731, 2017.

[35] O. Selezneva and D. Wolf, "Tools for Assuring WIM Data Quality: Practical Guide," National Cooperative Highway Research Program, 2023.

[36] ERA BMS-2025. Retrieved from https://bms.era.gov.et:8087/home/index (accessed on Jan. 15, 2025).

[37] Negarit Gazeta, "Vehicles Weights and Dimensions Regulations", FDRE-Council of Ministers, 2014.

[38] I. P. Michael, "Overload Control Practices in Eastern and Southern Africa, Main Lessons Learned," *Sub-Saharan Africa Transport Policy Program*, 2010.

[39] B. Sivakumar, "Legal Truck Loads and AASHTO Legal Loads for Posting," *Transportation Research Board*, vol. 575, 2007.

[40] FHWA, "Bridge Formula Weights" Federal Highway Administration, 2019.

[41] TTTFP, "Workshop on Weighbridge Best Practices," Ref. Europe Aid/138372/IH/SER/MULTI, RFS N. 2015/367563, 2019.

[42] R. R. A. Briones, "Bridge Load Rating: A General Procedure for Load Rating Bridges without Plans," PhD dissertation, Purdue University, 2018.

[43] JICA- Japan International Cooperation Agency, "Manual for Load Rating of Bridges," Department of Public Works and Highways, Republic of the Philippines, 2014.

[44] M. Ghosn, F. Moses, and J. Wang, "Redundancy in Highway Bridge Superstructures," FHWA Report No. FHWA-HRT-04-039, 2003.

[45] E.C. Bentz and M.P. Collins, "Response 2000: User Manual," 2001.

[46] EthioInfra, "Determination of Residual Bearing Capacity of Bridges along Selected Trunk Roads," Final Report-Lot-1: Alemgena, Nekempt, Jimma and Sodo Districts, 2021.

[47] I. Paeglite, A. Paeglitis, and J. Smirnovs, "Dynamic Amplification Factor for Bridges with Span Length from 10 To 35 meters,". *Engineering Structures and Technologies*, vol. 6, no. 4, 2014, pp.151-158.

[48] SDDoT, "Load Rating Manual," Office of Bridge Design, South Dakota Department of Transportation, 2024.

[49] R. Nick, "Unique Insight into the World of Bridge Engineering," https://thebridgeguy.org/2024/09/the-mystery-ofload-rating/.Accessed August 07, 2025.

[50] C. Wang and Q. Li, "Simplified Method for Time-Dependent Reliability Analysis of Aging Bridges Subjected to Nonstationary Loads," *International Journal of Reliability, Quality and Safety Engineering*, vol, 23, no, 1, 2016, pp. 1-14.

[51] F. Chehade, R. Younes, H. Mroueh, and F. H. Chehade, "Time-Dependent Reliability Analysis of Reinforced-Concrete Bridges under the Combined Effect of Corrosion, Creep and Shrinkage," *WIT Transactions on the Built Environment,* vol. 174, 2018, pp. 13-24.

[52] T. El Maaddawy, K. Soudki, and T. Topper, "Long-Term Performance of Corrosion-Damaged Reinforced Concrete Beams," *ACI Structural Journal*, vol. 102, no. 5, 2005, pp. 649-656.

[53] A. Gebre, B. Wesenseged, Y. Gebre, and G. Segni, "Live Load Models for Ethiopian Highway Bridges using Truck Weigh in Motion (WIM) Data," *Transportation Research Record*, vol. 2679, no. 10, 2025, pp. 1-18.

[54] C. D. Eamon, V. Kamjoo, and K. Shinki, "Design Live Load Factor Calibration for Michigan Highway Bridges," *Journal of Bridge Engineering*, vol. 21, 2016.

[55] B. Sivakumar, and M. Ghosn, "Recalibration of LRFR Live Load Factors in the AASHTO Manual for Bridge Evaluation," NCHRP Project 20-07, 2011.

[56] F. Moses, "Calibration of Load Factors for LRFR Bridge Evaluation," NCHRP Report-454, *Transportation Research Board*, Washington, DC, USA, 2001.

Experimental Study on the Efficiency of Passive Auto-Tuning Compound Pendulum Mass Damper

Yirga Terefe[1,*], Abrham Gebre[2]

[1]*School of Civil and Environmental Engineering, Ambo University, Ambo, Ethiopia.*
[2]*School of Civil and Environmental Engineering, College of Technology and Built Environment, Addis Ababa University, Addis Ababa, Ethiopia.*
**Corresponding author E-mail address: yirga.afomia@gmail.com*

DOI: https://doi.org/10.63990/zede.v43i.12970

ABSTRACT

A Tuned Mass Damper (TMD) is a device used to reduce the effects of dynamic responses of a structure during seismic action. In this study, a test model of a two-story steel structure was used to evaluate the efficiency of the Passive Auto-Tuning Compound Pendulum Mass Damper (PATCPMD). The PATCPMDs were suspended in the structure's top and lower stories and controlled by a group of flexible ropes that formed a compound pendulum, but it was not quite a compound pendulum and could move in any translational direction. The results showed that use of PATCPMD can provide significant control over the structure's translational, torsional, and coupled vibrations, with a maximum reduction in peak SSMS of 75 % for translational vibrations and up to 65 % for torsional vibrations when they are suspended in the first story. These values increased to 90 % for translational and 87 % for torsional vibrations when suspending in the second story. For forced vibrations, the maximum reductions in vibration control achieved were 68 % and 89 % if the damper was suspended in the first and second floor levels, respectively. Results showed that using PATCMD is more efficient when suspended on the second floor.

Keywords: Coupled vibrations, Steel structure, PATCPMD, Torsion, Translation,

1. INTRODUCTION

Modal analysis is a technique used in structural engineering to understand the dynamic behaviour of structures or systems. It involves the study of the natural frequencies, mode shapes, and damping properties of a structure or system under various conditions [1, 2]. In modal analysis, the words "participation" and "dominant modes" are important concepts concerning the behaviour of structures under dynamic loading situations. Participation refers to the extent to which a specific mode of vibration contributes to the overall response of a structure or system under dynamic loading, whereas dominant modes are those modes of vibration that contribute the most to the dynamic response of a structure under specific loading conditions [3].

In recent years, the development of active and passive control devices has become an interesting research area into vibration control of civil engineering structures. Passive control devices are those that are actuated by structural stimulation but do not provide feedback. They tend to be popular since they require little energy to operate and are mechanically simple and efficient. TMD is a prominent passive control device that is frequently used in buildings. The World Trade Centre Tower (1973) in New York and the John Hancock Tower (1976) in Boston were among the first practical examples of this type of control technology. Since then, a number of high-rise buildings, towers, bridges, chimneys and mast structures have been

equipped with TMD to suppress wind-induced vibrations [4].

A tuned mass damper from viscoelastic material was developed and discovered that it is effective when tuned to the natural frequency over a limited band. It was also discussed that how to estimate the viscous damping of a TMD composed of viscoelastic material. For any given floor mass, damping, and stiffness, a damper can be an inexpensive and simple solution for retrofitting floors with excessive vibrations [5]. Moreover, similar studies on application of TMD in framed structures were studied [6, 7] and effective results were found if TMD with less damping ratio were used [6]. In similar research, to determine the efficiency and parameters of a passive auto-tuning mass damper (PATMD), an experimental work was conducted. The results of the 'PATMD efficiency tests' reveal its capacity to provide significant control over the structure's translational, torsional, and coupled vibrations without being tuned in any way. The tests also demonstrate that the PATMD is robust, simple, and versatile, providing it an ideal application for engineering structures [8]. It is also reported that TMD can be used to control structural vibrations [9,10].

Similarly, experimental investigation was carried out using TMD with variable stiffness. The output of the experimental test revealed that the system controls the human-induced vibrations effectively and identifies the natural frequency accurately [11]. A research based on numerical analysis and experimental investigation was made for bridge structures with TMDs and good result was found without deforming the structure [12].

To investigate the structure's behaviour with and without TMD, a one-story and two-story building frame model were created for a shake table experiment under sinusoidal excitation. The TMD was tuned to the structural frequency while maintaining the stiffness and damping ratio constant. Various parameters, including frequency ratio, mass ratio, tuning ratio, and so on, were used to assess the TMD's efficiency and robustness in terms of percentage reduction in structure amplitude. The responses were then numerically validated using the finite element approach, demonstrating that TMD may be efficiently used to control structural vibration [13].

In other experimental works, the mass damper parameters were tuned using an evolutionary operation (EVOP) algorithm. To develop a computer program, the El Centro NS earthquake data was used and it was found that a higher percentage of reduction on the roof of a ten-story structure using TMD with EVOP algorithm [14].

Similarly, group of researchers investigated vibration control for seismic structures using semi-active friction multiple tuned mass dampers (SAF-MTMD). Various friction mechanisms were used to activate all of the mass units of the SAF-MTMD during an earthquake. The study showed that SAF-MTMD effectively minimizes seismic motion, particularly at higher intensities [15]. Some experimental work was also performed using a pendulum tuned mass damper with advantages over standard TMD and discovered that the frequency may be re-tuned by adjusting the cable length [16].

Active auto-tuning compound pendulum mass dampers have generally been effective in many applications; however, they have major drawbacks: sensitivity to detuning, more prone to failure as the actuator relies on electronic components, large power consumption, installation issues, and so on [17, 18]. Moreover, third-generation active dampers have drawbacks such as design complexity, installation and operation costs, the need for continuous power supply and technical challenges for high vibration amplitudes. On the other hand, PATCPMD is simple, effective, and

easily adaptable [17]. Hence, in this study, the efficiency of PATCPMD in regulating vibration and dampening responses of a steel floor system model was experimentally investigated.

2. METHODS

2.1. Materials

Mild steel was used for both the rods and the plates. The top and bottom plates weighed5.587kg and 5.692kg, respectively. The column dimensions were 9mm in diameter and each story has a height of 0.69m, with an overall structural height of 1.38m.

2.2. Equipment

2.2.1. Unidirectional shaking table

The unidirectional shaking table consisted of a 0.5m×0.5m sliding plate with a thickness of 6mm. The sliding surface had 81 holes on a 25×25mm grid and the shaking table had a maximum payload of 49.023kg. Figure 1 shows a shaking table that can operate at frequencies ranging from 0 to 6Hz, and controlled by a speed controller and a tachometer. It had the ability to produce simple harmonic motion with a maximum tilt of 20mm (± 5mm).

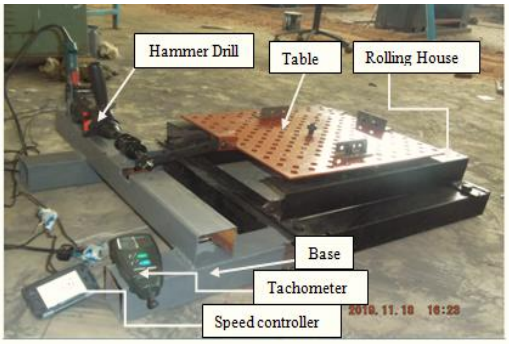
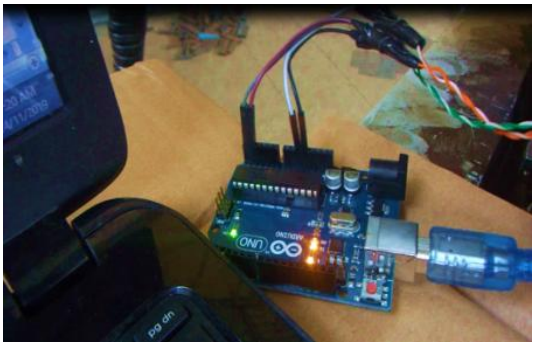


Figure 1 The shaking table

2.2.2. Gyro accelerometer and Arduino Uno board

A Gyro accelerometer shown in Figure 2(a) was attached to the frame at the location where the acceleration needs to be measured. The Gyro accelerometer was attached to an Arduino uno board (Figure 2(a)) via four data cables, which recorded acceleration and rotation data over time.



(a)

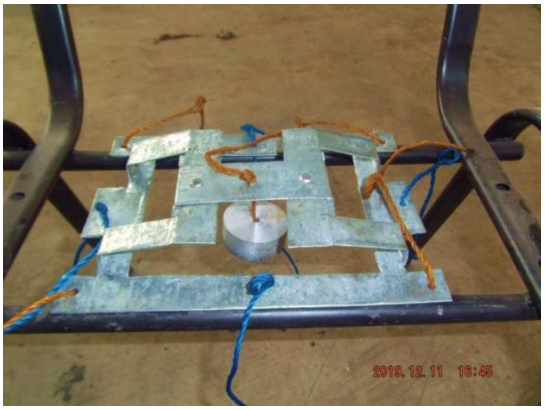
(b)

Figure 2 (a) Arduino Uno board (b) Gyro accelerometer attached to the top story.

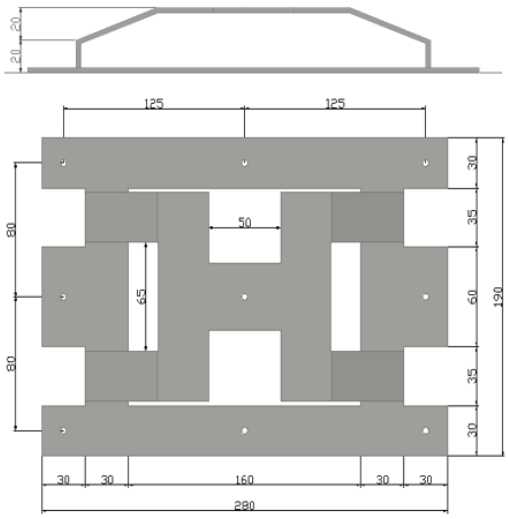
2.2.3. PATCPMD

The PATCPMD shown in Figure 3was manufactured as a compound pendulum, although not exactly, with the link made of galvanized sheet metal weighing 0.457kg and the damper was made of aluminum weighing 0.153kg. The damper had a total mass of 0.611kg, which was approximately 4 % of the basic model's mass of 14.104kg.

The compound pendulum had a hole in the top center to suspend the mass damper, and four holes on the base to support the entire pendulum on the model's floors. The damper was suspended from the four holes at a perpendicular height of 70mm below the plates by inelastic ropes, with the damper mass suspended in the middle of the link at a height of 40mm. The steel structure model and PATCPMD suspended at the second story of the structure is shown in Figure 4.



(a)



(b)

Figure 3 PATCPMD (a) photo and (b) schematic drawing

(a) (b)

Figure4 (a) Test model (b) PATCPMD suspended at the second story

2.2.4. Measuring Tape, Wrenches, Dial Gauge and Tachometer

The tachometer shown in Figure 5was used to measure the speed of the drill which in turn controlled the oscillation of the shaking table. In addition to the tachometer, measuring tape, and dial gauge were used.

Figure 5 Tachometer

2.2.5. Speed controller

The speed controller was managed by an input voltage of 220 volts. A dimmer switch was used to controlthe speed of the hammer drill machine, while a tachometer was employed for controlling the drilling machine's rotation and also the shaking table's excitation frequency in the range 0 to 6Hz.

2.3. Experimental Setup

Figure 6 shows the experimental setup, which includes a unidirectional shaking table, gyro accelerometer and Arduino Uno board, speed controller, measuring tape, wrenches, dial gauge, tachometer, and PATCPMD. In the experimental setup, the damper is suspended at different floor levels of the steel-structured test model.

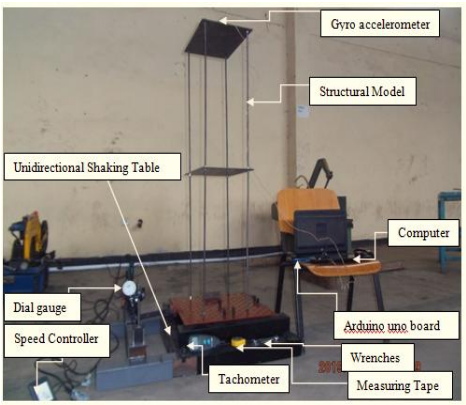


Figure 6 Experimental setup

The main structural configuration shown in Figure 6 consisted of four columns made of 1.38m high, 9mm diameter threaded steel rods, and two slabs with 300mm × 400mm × 6mm thick steel plates. The plates were fixed at heights of 0.69m and 1.38m for the first and second stories, respectively, with nuts. Each plate has four holes drilled for suspending the PATCPMD to the model. The columns were securely fastened to a unidirectional shaking table platform, ensuring that the ends were rotationally fixed.

2.3.1. Experimental Procedure

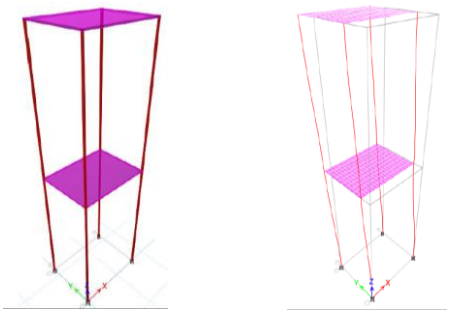
The effectiveness of PATCPMD in suppressing transitional, tensional, and coupled vibrations was examined. The free and forced vibration testing were carried out during the PATCPMD efficiency tests. Both tests were performed on the primary test model, with and without the PATCPMD. The type of tests is summarized in Table 1. The fundamental frequencies of the structure were determined via free vibration analysis. The time history of acceleration analysis plots provides information to describe the structure's dynamic behavior. Frequency analysis provides useful information regarding structural vibration. Time-history signal can be transformed to the frequency domain. The Fourier Transform (FT) is the most commonly used mathematical approach for transforming time signals into frequency domains. In structural analysis, time wave-forms are often measured and their Fourier Transforms developed. In this research, to get the time-history responses, a Fast Fourier Transform (FFT) MATLAB code was used.

Table 1 Tests undertaken for PATCPMD efficiency

No.	Type of tests	Remarks
1	Free translational vibration	The model was excited by applying an initial displacement of 10mm, measured with a dial indicator, and then releasing it.
1.1	Translational vibration in x-direction	
1.2	Translational vibration in y-direction	
2	Free torsional vibration	The model was twisted at the top story until it reached 10mm of translational displacement before being released.
3	Coupled vibration (translational and torsion)	Before being released, the model was subjected to 10mm of translational and torsional movement at the top story.
3.1	Coupled vibration in x-direction	
3.2	Coupled vibration in y-direction	
4	Forced vibration	The test model was first oriented on the shaking table to generate forced vibrations in the x-direction, and then it was turned 90° to experience forced vibrations in the y-direction.
4.1	Translational vibration in x-direction	
4.2	Translational vibration in y-direction	

2.4. Validation

Extended Three-Dimensional Analysis of Building System (ETABS) structural software was used to simulate the behavior of the primary model, which also determined its first natural frequencies and associated mode shapes. The sample model used for validation and corresponding translation in y-direction is shown in Figure 7.The dimensions of the model given in section 2.3 were entered into the ETABS software and used to validate the proposed experimental procedure and test results. The model's natural frequency for translation in both x- and y-directions and a comparison to experimental results are shown in Table 2.



a) b)

Figure 7. a) Frame model in ETABS and b) deformed shape

As shown in Table 2, the analytical results are 92 % accurate with the observed experimental values, showing that the results obtained from the experiment are found to be valid. It implies that the numerical modeling in ETABS is also considered to account for uncertainties in experimental setups [13].

Table 2 Natural frequency for free vibration

Direction	Experiment (1)	Analysis (2)	Ratio (1)/(2)
x	3.0151Hz	3.297Hz	0.916
y	3.1156Hz	3.355Hz	0.928

3. RESULTS AND DISCUSSION

This study was conducted to assess the efficiency of PATCPMD, and it presents several novel contributions that set it apart from previous research in the field. Some of these include:

i) *Innovative design concept*: the study introduces a new design concept of a PATCPMD.

ii) *Auto-tuning mechanism*: the research explores the integration of an auto-tuning mechanism within the compound pendulum mass damper.

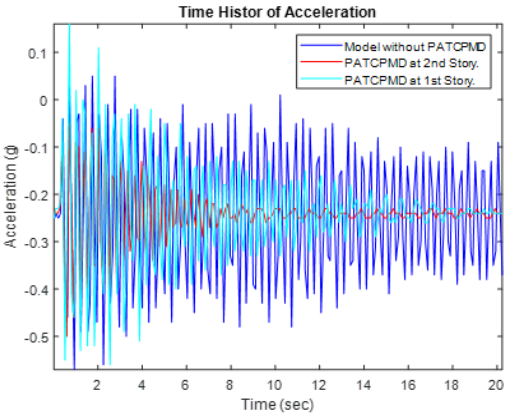
iii) *Efficiency*: an experimental work for evaluating the efficiency of the PATCPMD is done and validated.

iv) *Practicability:* the research underlines the simplicity and practical significance of the PATCPMD in mitigating structural vibrations.

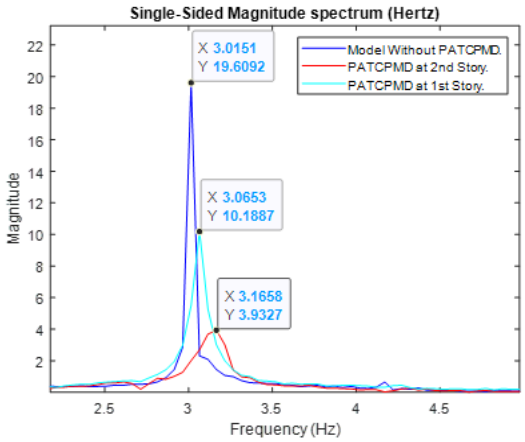
Thus, in the next sub-sections the efficiency of the proposed PATCPMD system is evaluated and discussed in detail.

3.1. Free Translational Vibration

In Figures 8 and 9, the time-history graph of the acceleration response and the single-sided magnitude spectrum (SSMS) acceleration response in the x- and y-directions are shown, respectively. The results are also summarized in Table 3.

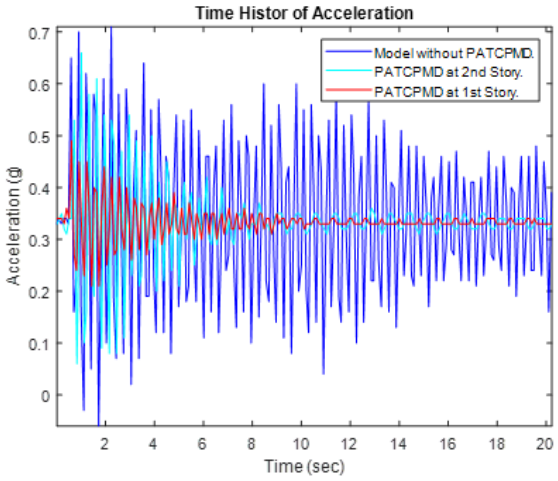


(a)

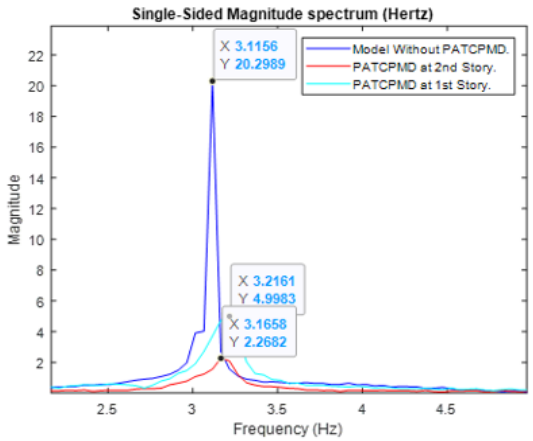


(b)

Figure 8 Free translational vibration responses in the x-direction (a) time history and (b) SSMS



(a)



(b)

Figure 9 Free translational vibration responses in the y-direction (a) time history and (b) SSMS

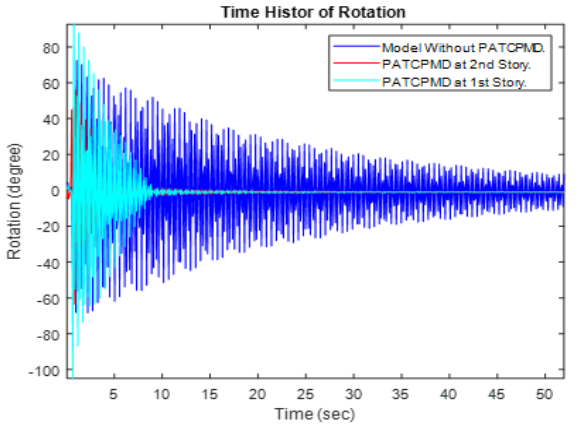
In Table 3, it is shown that the PATCPMD provides a high level of peak attenuation when hanging at the second story. The peak frequency magnitude in the x-direction, 19.609, reduced by 48 % and 80 % when the PATCPMD was placed at the first and second stories, respectively, while the peak frequency magnitude for translation in the y-direction, 20.299, is reduced by 75 % and 89 % when the PATCPMD was suspended on the first and second stories, respectively. The experimental results of this study are larger than related studies, where a 57.8 % reduction in vibration responses of the system was achieved with the application of tuned mass dampers [11].

Table 3 Free translational vibration results

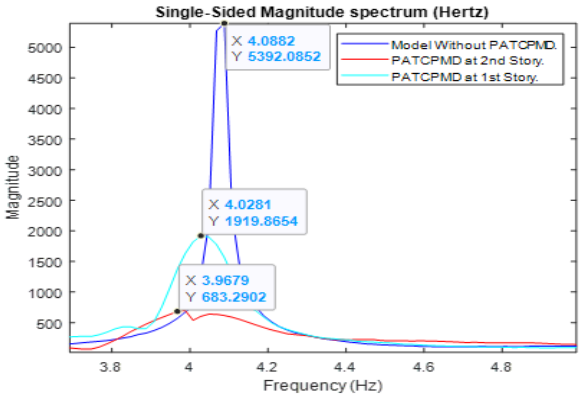
Direction	without PATCPMD	PATCPMD on 1st floor	PATCPMD on 2nd floor
Natural frequency magnitude (Hz)			
x	3.015	3.065	3.166
y	3.116	3.216	3.166
Attenuation in the translational direction			
x	19.609	10.189	3.933
y	20.299	4.998	2.268

3.2. Free Torsional Vibration

In a free torsional vibration, the system experiences torsional (twisting) motion due to the elastic properties of the materials involved [16]. The time-history and SSMS rotation responses for the free torsional vibration are shown in Figures 10(a) and (b), respectively.



(a)



(b)

Figure 10 Free torsional vibration responses (a) time history (b) SSMS

In Table 4, summary of the model's free torsional vibration results is given.

Table 4 Free torsional vibration results

without PATCPMD	PATCPMD on 1st floor	PATCPMD on 2nd floor
Natural frequency magnitude (Hz)		
4.088	4.028	3.968
Attenuation in the translational direction		
5,392.085	1,919.866	683.290

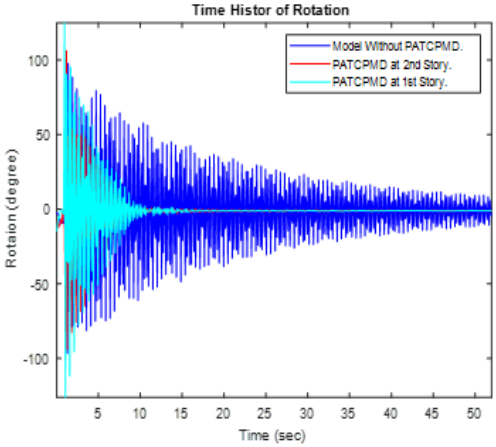
As shown in Table 4, the use of PATCPMD reduced the peak torsional frequency by 64.4 % and 87.3 % when it was suspended at the first and second floors, respectively. These values are 6.60 % and 29.5 % higher than those obtained in a previous study [11], which reported a 57.8 % reduction in vibration amplitude.

3.3. Coupled Vibration (Translation and Torsion)

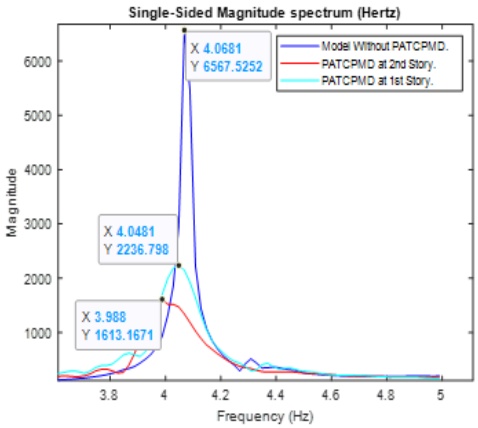
Coupled vibration, involving both translation and torsion, occurs when a structure experiences simultaneous movement in both linear (translational) and rotational (torsional) directions [19].

3.3.1. Coupled vibration in x-direction

In Figures 11(a) and (b), the time-history and SSMS rotation responses of the structure due to coupled vibration in x-direction are shown, respectively. Similarly, the acceleration responses are shown in Figure 12. As shown in these figures, suspending PATCPMD in the second story reduced the structure's SSMS rotation by 75 % and acceleration by 85 % when compared to the normal model. A 66 % reduction in SSMS rotation was achieved when the damper was mounted in the first story. The results are comparable with similar research suggesting that the use of TMDs is substantially more successful at reducing structural vibration by up to 77.28 % when subjected to sinusoidal ground excitations [20].

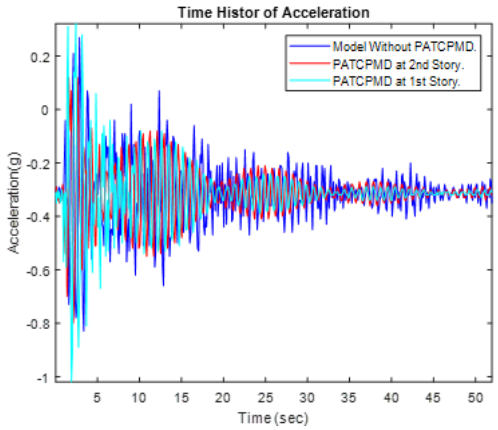


(a)

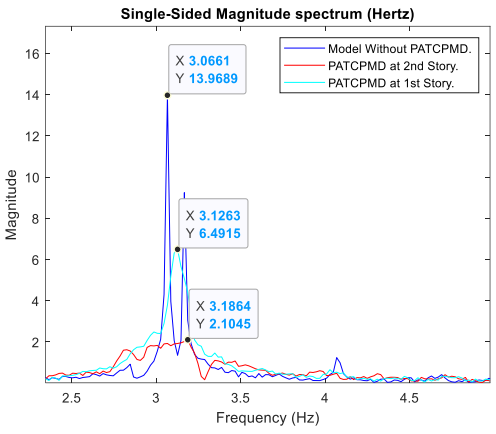


(b)

Figure 11 Coupled torsional responses in x-axis (a) time-history and (b) SSMS



(a)

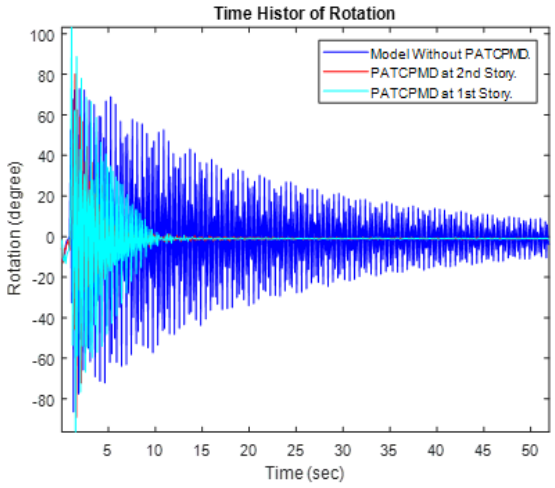


(b)

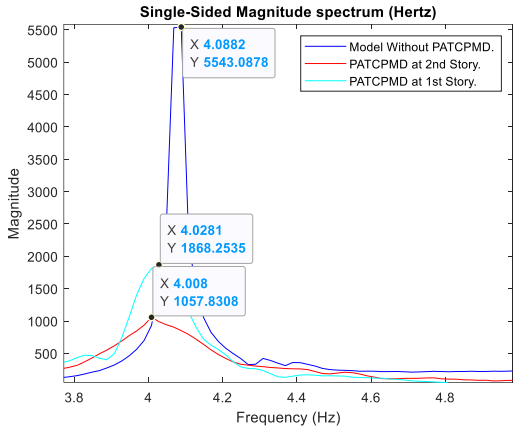
Figure 12 Acceleration responses in x-direction (a) time-history and (b) SSMS

3.3.2. Coupled vibration in y-direction

Figures 13 and 14 show the time history rotation and SSMS acceleration responses for both rotation and acceleration.

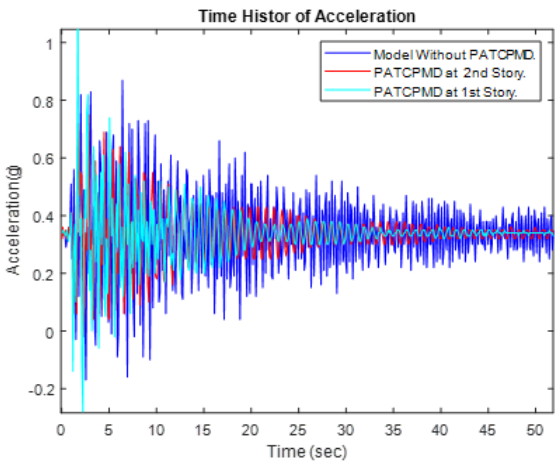


(a)

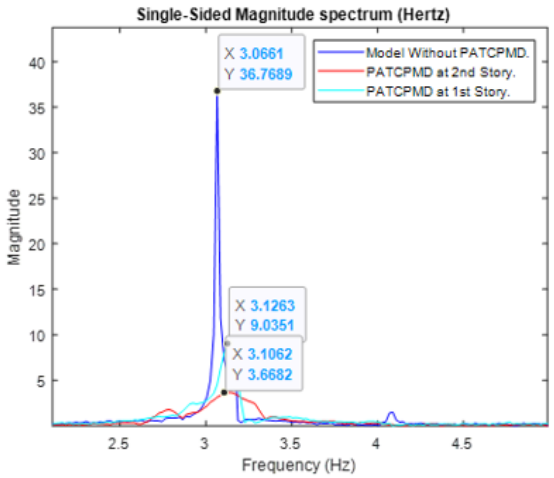


(b)

Figure 13 Coupled torsional responses in y-axis (a) time history and (b) SSMS



(a)



(b)

Figure 14 Acceleration responses in y-direction time-history and (b) SSMS

Summary of rotation and acceleration responses of the model in x- and y-directions due to coupled vibration is shown in Table 5. As indicated in the table, when the PATCPMD is suspended at the first and second stories with forced vibration in the y-direction, the peak rotation frequency magnitude reduced by 66.3 % and 80.92 %, respectively. Whereas the peak acceleration frequency magnitude reduced by 75.43 % and 90 % when the damper was suspended in the first and second stories, respectively. The results are comparable with previous studies that reported a 58 % reduction [11].

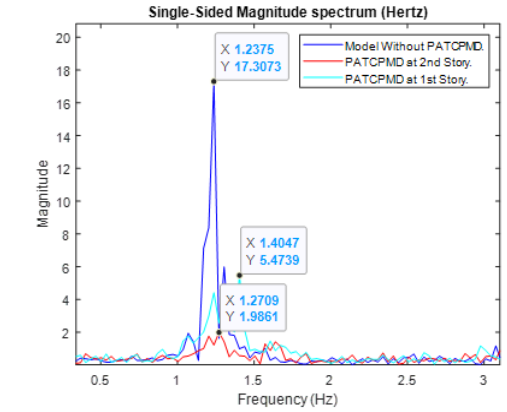
Table 5 Coupled vibration test results

Responses	without PATCPMD	PATCPMD on 1st floor	PATCPMD on 2nd floor
x-direction			
rotation	6,567.52	2,236.79	1,613.16
acc.	13.97	6.492	2.105
y-direction			
rotation	5,543.09	1,868.25	1,057.83
acc.	36.769	9.035	3.668

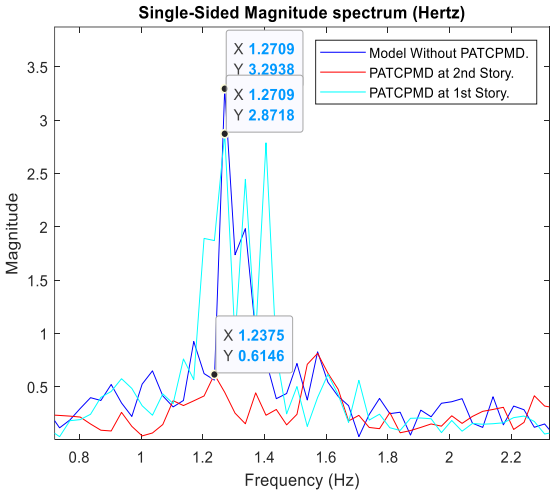
3.4. Forced Vibration: Translational Vibration in x- and y-directions

The test model was initially oriented on the shaking table to generate forced vibrations in the x-direction, and then it was rotated by 90° to experience forced

vibrations in the y-direction. The SSMS acceleration responses are shown in Figure 15 and the results are summarized in Table 6.



a) x- direction



b) y- direction

Figure 15 SSMS of the acceleration responses for forced translational vibration

Table 6 Forced translational vibrations in x- and y-directions

Direction of forced vibration	Natural frequency magnitude		
	without PATCPMD	PATCPMD on 1st floor	PATCPMD on 2nd floor
x	17.307	5.474	1.986
y	3.294	2.872	0.615

The results shown in Table 6indicate that the SSMS acceleration response is peak at 17.307 and 3.294 in the x and y directions, respectively. The PATCPMD's obviously controlled the forced vibration tests, where the peak vibration in the x-direction reduced by 88.5 % and 68.4 % when the PATCPMD was suspended at the second and first story, respectively. Moreover, in the y-direction, the values reduced by 81.3 % and 13 % when it was suspended at the second and first story, respectively. When the damper was suspended at the first story, the result was comparable to those published in [21], which showed that TMD reduced structural vibration by at least 13 %.

Generally, according to the experimental results of this study, in all type of tests, the frequencies ranged between 0.615 Hz and 5.474 Hz. Code recommends that structures with lowest natural frequency above 1 Hz is small and the resonant response can be ignored [22].The period of the structure for dampers with a 5 % damping ratio is between$0.2T_1$ and $2T_1$, where T_1 is the fundamental period of the structure [23]. In this study, the fundamental frequency of the frame (*f*) was3.166 Hz (from free vibration analysis, Table 2), and accordingly, the fundamental period of the structure was found to be 0.321sec (T_1=1/3.166). As a result, the structure's period ranged from 1/5.474 to 1/0.615 (0.182 sec to 1.626 sec); in some cases, the limit is exceeded ($0.2T_1$ to $2T_1$; 0.064 sec to 0.642 sec).

4. CONCLUSIONS

The vibration attenuation device proposed in this study, the PATCPMD, addresses the issues of excessive torsional and coupled vibrations.

The results of the experimental work indicate that the SSMS translation and rotation responses reduced by 48 % to 75 % when the PATCPMD was suspended in the first story. These reductions increased to 90 % if the damper was suspended in the second story. Furthermore, for forced vibrations, the reduction varied from 13 % to 68 % and 75 % to 80 %, when the PATCPMD was suspended at first and second stories respectively, indicating that PATCPMD was able to effectively control

all types of excitations. Effective control of excitation was recorded when the PATCPMD is suspended in the second story as compared to placing it at the first story.

The effectiveness of PATCPMD in a multi-story building should be examined and further research is needed to assess the efficiency of PATCPMD under earthquake loading and frequency sweep using various parameters.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Mr. Tekleye Geda and other laboratory staff from Civil, Mechanical, and Electrical Engineering Department at Ambo University, Institute of Technology for their assistance in developing the test model for this study.

REFERENCES

[1] P. Avitabile, “Experimental Modal Analysis.” *Sound and Vibration*, 35(1), pp.20-31, 2001.

[2] B. J. Schwarz, and M. H. Richardson, “Experimental Modal Analysis”, CSI Reliability Week, 35(1), pp.1-12, 1999.

[3] D. J. Ewins, “Modal Testing: Theory, Practice and Application”, John Wiley & Sons, 2009.

[4] M. Carlisle, “Experimental Investigations into a Passive Auto-Tuning Mass Damper”, (MSc. Thesis), University of the Witwatersrand, Faculty of Civil and Environmental Engineering, July 2013.

[5] I. Saidi, A. D. Mohammed, E. F. Gad, et al., “Optimum Design for Passive Tuned Mass Dampers Using Viscoelastic Materials”, *Aust. Earthq. Eng. Soc.* 2007 Conf., pp. 1–8, 2007.

[6] R. Mishra, "Application of Tuned Mass Damper for Vibration Control of Frame Structures under Seismic Excitations", (MSc. Thesis), National Institute of Technology, Rourkela, 2011.

[7] M. Yuan, A. Sadhu, and K. Liu, "Condition Assessment of Structure with Tuned Mass Damper using Empirical Wavelet Transform", *Journal of Vibration and Control, 24*(20), pp.4850-4867, 2018.

[8] E. N. Naicker, and K. Li, “Experimental Investigation on a Passive Auto-Tuning Mass Damper for Vibration Control”, *Int. J. Dynamics and Control,* vol. 6, no. 3, pp. 1047–1062, 2018, doi: 10.1007/s40435-017-0381-z.

[9] F. Rahimi, R. Aghayari, and B. Samali, "Application of Tuned Mass Dampers for Structural Vibration Control: A State-of-the-Art Review", *Civil Engineering Journal*, pp.1622-1651, 2020.

[10] Elias, S. and Matsagar, V., "*Research Developments in Vibration Control of Structures using Passive Tuned Mass Dampers*", *Annual Reviews in Control,* 44, pp.129-156, 2017.

[11] Wang, L., Nagarajaiah, S., Zhou, Y. and Shi, W., "Experimental Study on Adaptive-Passive Tuned Mass Damper with Variable Stiffness for Vertical Human-Induced Vibration Control", *Engineering Structures,* 280, p.115714, 2023.

[12] Nadu, T., "Experimental Investigation of Vibration Suppression for Avoiding Bridge Collapse by Pendulum Type Passive Tuned Mass Damping System", *Int. J. Eng. Res. Technol,* 9, pp.1156-1162, 2020.

[13] S. Padmabati, “*Experimental and Numerical Study on Tuned Mass Damper in Controlling Vibration of*

Frame Structures", No. 213, p. 59, 2015, http://ethesis.nitrkl.ac.in/7605/

[14] B. Islam, "Optimization of Tuned Mass Damper Parameters using Evolutionary Operation Algorithm", 15th World Conf. Earthquake. Eng., no. 1998, 2012.

[15] C. C. Lin, L. Y. Lu, G. L. Lin, and T. W. Yang, "Vibration Control of Seismic Structures using Semi-Active Friction Multiple Tuned Mass Dampers", *Eng. Struct.,* vol. 32, no. 10, pp. 3404–3417, 2010, doi: 10.1016/j.engstruct.2010.07.014.

[16] S. Chatterton, L. Dassi, E. Gheller, T. Ghisi, et al. "Torsional Vibration Analysis Using Rotational Laser Vibrometers", *Sensors,* 24, 1788, 2024. https://doi.org/10.3390/s24061788

[17] J. Connor, S. Laflamme, J. Connor, and S. Laflamme, "Tuned Mass Damper Systems", *Structural Motion Engineering,* pp.199-278, 2014.

[18] L. Koutsoloukas, N. Nikitas, and P. Aristidou, "Passive, Semi-Active, Active and Hybrid Mass Dampers: A Literature Review with Associated Applications on Building-Like Structures", *Developments in the Built Environment,* 12, 2022.

[19] Q. Zhang, C. Li, J. Zhang, and J. Jin, "Active Vibration Control and Coupled Vibration Analysis of a Parallel Manipulator with Multiple Flexible Links." *Shock and Vibration,* 2016. http://dx.doi.org/10.1155/2016/7474085

[20] P. Kundu, "Vibration Control of Frame Structure using Multiple Tuned Mass Dampers", (MSc. Thesis), 2012.

[21] V. M. Thakur, and P. D. Pachpor, "Seismic Analysis of Multistoried Building with TMD (Tuned Mass Damper)", *International Journal of Engineering Research and Applications (IJERA),* 2, pp.319-326, 2012.

[22] SEI/ASCE 7-02, "ASCE Standard-Minimum Design Loads for Buildings and other Structures", *American Society of Civil Engineers,* 2002.

[23] ES-EN-1998: 2015, "Design of Structures for Earthquake Resistance - Part 1: General Rules - Seismic Actions and Rules for Buildings", 2015.

Base Isolation for Earthquake Protection of Structures Considering Ethiopian Standard

Daniel H. Zelleke[1, *] and Vasant A. Matsagar[2]

[1, *] *Department of Civil Engineering, Haramaya Institute of Technology, Haramaya University, Haramaya, Ethiopia.*
[2] *Department of Civil Engineering, Indian Institute of Technology- Delhi, India.*
[*] *Corresponding Author's E-mail Address: E-mail address: daniel.habtamu@alumni.iitd.ac.in*

DOI: https://doi.org/10.63990/zede.v43i.12971

ABSTRACT

An increase in construction activities and recent advances in structural engineering necessitated the major revision of the Ethiopian Building Code Standard (EBCS), published in 1995. *The updated building code has introduced improved earthquake-resistant design considerations, including provisions for base-isolated (BI) structures. This study investigates the efficacy of the base isolation technique in earthquake protection of buildings considering Ethiopian standard. Moreover, the validity of the specific provisions of the Ethiopian seismic standard, i.e., Ethiopian Standard European Norm, (ES EN* 1998-1:2015) *on the choice of base isolator properties for analysis and design is investigated. Non-linear dynamic response history analyses of multi-story BI buildings are performed under synthetic earthquakes, matching with the response spectrum of the Ethiopian standard. Furthermore, the vibration response of fixed-base building models is reported for comparison. Four structural response quantities, i.e., the floor acceleration, base shear, inter-story drift, and isolator displacement, are studied. The findings demonstrate that the application of the base isolation technique reduces the dynamic response of multi-story buildings substantially. In addition, it is shown that some of the Ethiopian seismic standard provisions on isolator parameter consideration are not in logical agreement with the earthquake behavior of BI buildings observed in the current study.*

Keywords: Base isolation; Design response spectrum, Earthquake; EBCS, Ethiopian standard.

1. INTRODUCTION

The design and construction of structures in Ethiopia were governed by a building code issued in 1995, namely the Ethiopian Building Code Standard (EBCS), up to 2015. Since 1995, a vast amount of knowledge has come into picture in areas of civil engineering. Notable recent advances in the area include the improvement of knowledge about earthquakes and their effect on structures. These advances resulted in the revision of many international building codes. Our awareness of earthquakes around the East African region has also increased significantly [1, 2]. The presence of the East African Rift and data gained from the recent seismic events in the region has rendered the zone earthquake-prone. The 1983 Hawassa, 1985 Langano, 1989 Afar, 2002 Mekelle, 2009 Ankober, 2010 Hosanna, 2011 Yirgalem, 2011 Jinka, 2014 Asayta, and 2016 Hawassa earthquakes, which damaged buildings and injured many, are noteworthy. In light of this, researchers [3 - 5] raised their concerns regarding the adequacy of the seismic provisions stipulated in EBCS 8, 1995 [6] in accounting for the effect of earthquakes in the region. Furthermore, the researchers recommended the revision of the seismic design standard. It is also advocated that an updated design earthquake loading be introduced in the revised building code. For instance, it is suggested to: (a) use a return period of 475

years instead of the 100 years return period adopted in EBCS 8, (b) implement the recent worldwide seismicity data (map) published by the Global Seismic Hazard Assessment Program, GSHAP, and (c) adopt the new six site class system to account for site amplification effects properly [3 - 5].

Due to the concerns raised in light of the recent technological advances in structural engineering, the Ethiopian building standard has been revised. Several earthquakes have also occurred in Ethiopia since the revision of the Ethiopian seismic design standard. The latest notable seismic events include 18 earthquakes with magnitudes of more than five that occurred in various parts of Ethiopia between January 2024 and March 2025 [7]. Some of these frequent events have caused substantial damage to structures, which further justifies the efforts to install enhanced seismic provisions in the relevant seismic design standard.

A notable enhancement in the revised Ethiopian seismic design standard, i.e., Ethiopian Standard European Norm, (ES EN 1998-1:2015 2015) [8] is the inclusion of provisions for the protection of structures using base isolation. Base isolation is a structural response control technique that introduces very high lateral flexibility and shifts the fundamental frequency of the building away from the predominant frequencies of seismic ground motions. This shift in the vibration period of the structure significantly reduces the earthquake energy transmitted to the superstructure [9]. It is established that the technique enhances the protection of primary structural members and secondary systems (e.g., non-structural elements) of buildings from damage, especially during large earthquake events [10 - 14]. The base isolation technique has been successfully implemented in several countries, such as the USA, Japan, India, New Zealand, Yugoslavia, and South Africa. Worldwide, base isolation has been used for high importance structures, such as bridges, industrial structures, hospitals, computer service (internet data) centers, and nuclear power plants. Also, the technique was successfully implemented on various types of building projects, such as hotels, offices, condominiums, schools, and dormitories [15 - 16]. Observed records from the real-life implementations have demonstrated that buildings equipped with the technique have shown excellent performance during earthquake events [15].

Although design provisions are included in the revised Ethiopian seismic standard, i.e., ES EN 1998-1:2015, for structures equipped with seismic isolation devices, the base isolation technique is not yet implemented for any structure in Ethiopia. Moreover, there are no studies that investigate the provisions of the standard in contrast to the behavior of base-isolated (BI) systems. Therefore, it is interesting and useful to study the structural merits of the base isolation technique in protecting important structures against the undesirable effects of seismic activity in the Ethiopian context. Moreover, it is also useful to study the provisions of the Ethiopian standard on base isolation. The objectives of this study include: (a) to assess the efficacy of base isolation technique in mitigating the earthquake-induced vibration of multi-story buildings considering the Ethiopian standard; (b) to explore the influence of the properties of three seismic isolation devices (i.e., laminated rubber bearing, LRB; lead-core rubber bearing, LCRB; and friction pendulum system, FPS) on the behavior of BI buildings; and (c) to study the validity of the Ethiopian seismic standard provisions on the consideration of base isolation device parameters for response assessment.

2. MATERIALS AND METHODS

2.1. Mathematical Modeling of Base-Isolated Building

The schematic representation of the multi-story BI building is presented in Figure 1, wherein the placement of the base isolators is depicted. Moreover, the three-dimensional (3D) schematic diagrams and idealized models of the three types of base isolators (i.e., LRB, LCRB, and FPS) used in this investigation are shown in Figures 2(a) and 2(b), respectively. The reasons for the selection of these three types of isolation systems are threefold. First, these three types of base isolation devices are among the most popular choices in the practice. Second, they represent the two primary categories of isolators (i.e., elastomeric (rubber-based) and sliding types of isolation systems). Third, their characteristics encompass a range of linear and nonlinear force-deformation characteristics. In this study, equal seismic masses are considered at all floor and base slab levels, whereas all stories are assigned the same lateral stiffness. Here, the superstructure of the building is assigned a modal damping ratio (ξ_s) of 0.02. The stiffness of the stories and the floor masses are decided to achieve the desired fixed-base (FB) fundamental time period (T_s).

For both the FB and BI buildings, the equations of motion are derived considering earthquake excitation. Here, the derived equations of motion of the BI building are arranged in the state-space form as given in Equation (1) [17].

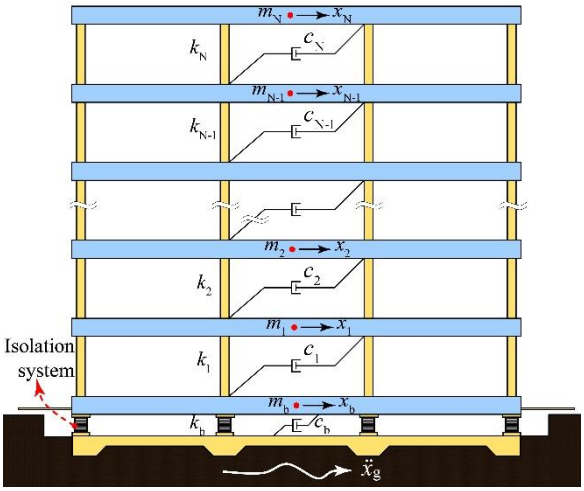


Figure 1 The schematic representation of the multi-story BI building.

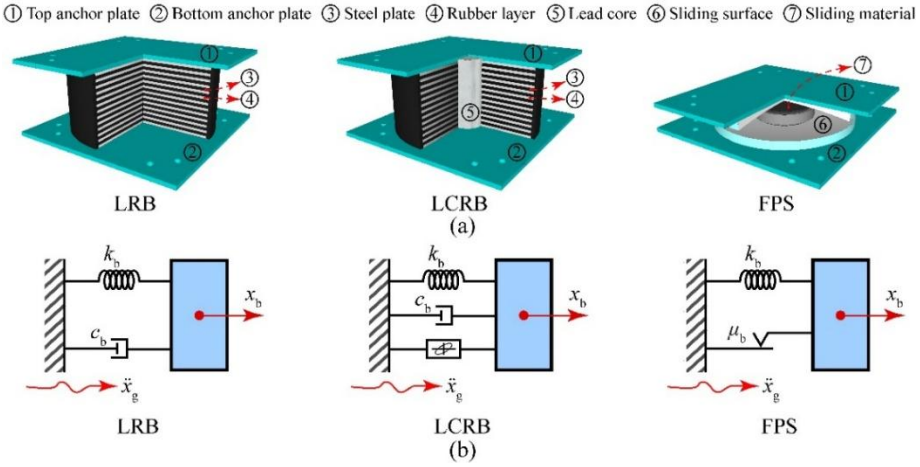


Figure 2 Details of LRB, LCRB, and FPS: (a) 3D schematic diagrams and (b) idealized mathematical models.

$$\begin{Bmatrix}\dot{\bar{X}}\\ \ddot{\bar{X}}\end{Bmatrix}=\begin{bmatrix}\vec{\mathbf{0}} & \boldsymbol{I}\\ -\bar{\boldsymbol{M}}^{-1}\bar{\boldsymbol{K}} & -\bar{\boldsymbol{M}}^{-1}\bar{\boldsymbol{C}}\end{bmatrix}\begin{Bmatrix}\bar{X}\\ \dot{\bar{X}}\end{Bmatrix}+\begin{bmatrix}\vec{\mathbf{0}}\\ -\bar{\boldsymbol{M}}^{-1}\end{bmatrix}\begin{Bmatrix}-1\\ \mathbf{0}^{\mathrm{T}}\end{Bmatrix}f_{\mathrm{b}}-\begin{bmatrix}\vec{\mathbf{0}}\\ -\bar{\boldsymbol{M}}^{-1}\end{bmatrix}\bar{\boldsymbol{M}}\begin{Bmatrix}-\ddot{x}_{\mathrm{g}}\\ -\boldsymbol{r}\left(\ddot{x}_{\mathrm{g}}+\ddot{x}_{\mathrm{b}}\right)\end{Bmatrix}, \tag{1}$$

where the mass, damping, and stiffness matrices, i.e., $\bar{\boldsymbol{M}}$, $\bar{\boldsymbol{C}}$, and $\bar{\boldsymbol{K}}$, of the BI building are defined in Equations (2), (3), and (4), respectively; $\bar{\boldsymbol{X}}=\{x_b \quad \boldsymbol{X}_s^T\}^T$ is the vector of structural displacements; f_b is the force in the seismic isolator; $\ddot{x}_g$ is the ground acceleration induced by earthquake; x_b and $\ddot{x}_b$, respectively, are the lateral displacement and acceleration response quantities of the base slab relative to the ground; $\boldsymbol{X}_s=\{x_1,x_2,...,x_N\}^T$ is a vector containing lateral floor displacements measured relative to the base slab; $\boldsymbol{I}$ is an identity matrix of size N+1; N represents the number of floors/stories in the structure; $\boldsymbol{0}$ is a null column vector; $\boldsymbol{r}=\{1,1,...,1\}^T$ is a column vector of influence coefficients; and $\overleftrightarrow{\boldsymbol{0}}$ is a null matrix.

$$\bar{\boldsymbol{M}}=\begin{bmatrix} m_b & \boldsymbol{0} \\ \boldsymbol{0}^T & \boldsymbol{M}_s \end{bmatrix}, \tag{2}$$

$$\bar{\boldsymbol{C}}=\begin{bmatrix} 0 & c_1\boldsymbol{r}^b \\ \boldsymbol{0}^T & \boldsymbol{C}_s \end{bmatrix}, \tag{3}$$

$$\bar{\boldsymbol{K}}=\begin{bmatrix} 0 & k_1\boldsymbol{r}^b \\ \boldsymbol{0}^T & \boldsymbol{K}_s \end{bmatrix}. \tag{4}$$

Here, m_b = mass of the base slab; c_1 = dashpot constant of the building's first story; k_1 = stiffness of the first story; $\boldsymbol{r}^b=\{-1,0,0,...,0\}$ is a row vector of size N; and $\boldsymbol{M}_s$, $\boldsymbol{C}_s$, and $\boldsymbol{K}_s$, respectively, are the superstructure mass, damping, and stiffness matrices. Moreover, the compact form of Equation (1) can also be written as $\dot{\bar{\boldsymbol{z}}}=\bar{\boldsymbol{A}}\bar{\boldsymbol{z}}+\bar{\boldsymbol{B}}\bar{\boldsymbol{F}}_c-\bar{\boldsymbol{B}}\bar{\boldsymbol{F}}_{exc}$, where $\bar{\boldsymbol{z}}=\{\bar{\boldsymbol{X}} \quad \dot{\bar{\boldsymbol{X}}}\}^T$ is the state vector. Also, the excitation vector ($\bar{\boldsymbol{F}}_{exc}$) and $\bar{\boldsymbol{F}}_c$ are given as

$$\bar{\boldsymbol{F}}_{exc}=-\bar{\boldsymbol{M}}\left\{\ddot{x}_g \quad \boldsymbol{r}\left(\ddot{x}_g+\ddot{x}_b\right)\right\}^T, \tag{5}$$

$$\bar{\boldsymbol{F}}_c=\{-1 \quad \boldsymbol{0}^T\}^T f_b. \tag{6}$$

The state-space solution [18], i.e., given in Equation (7), is implemented to quantify vibration response of the BI buildings under seismic excitation.

$$\bar{\boldsymbol{z}}_t=e^{\bar{\boldsymbol{A}}\Delta t}\bar{\boldsymbol{z}}_{t-\Delta t}+e^{\bar{\boldsymbol{A}}\Delta t}\Delta t\bar{\boldsymbol{B}}(\bar{\boldsymbol{F}}_c-\bar{\boldsymbol{F}}_{exc}), \tag{7}$$

where $\bar{\boldsymbol{z}}_t$ represents the state vector at time instant t.

The force in the LRB (f_b), given in Equation (8), is quantified as a function of the dashpot constant (c_b) and the lateral isolator stiffness (k_b). The damping coefficient is quantified as $c_b=2\xi_b M\omega_b$, where ξ_b, $\omega_b=2\pi/T_b$, T_b, and M are the damping ratio of isolation system, isolation angular frequency, isolation time period, and the total mass of the BI building, respectively. Further, the isolation stiffness is evaluated as $k_b=M\omega_b^2$. The force in the LCRB (f_b) is defined in Equation (9) [19] in terms of the damping coefficient, lateral stiffness, post-yield stiffness ratio ($\alpha=k_b/k_i$), and normalized yield strength (F_0). Here, the normalized yield strength can be obtained as $F_0=F_y/W$, where F_y and W, respectively, are the isolator yield strength and the weight of the BI building. Further, the initial stiffness is defined as $k_i=F_y/q$, where q = bearing yield displacement. The force in the FPS is given in Equation (10) as a function of the coefficient of friction of the FPS (μ_b) and isolator stiffness (k_b). For the FPS, the isolation stiffness is computed as $k_b=W/r_b$, where r_b is the radius of curvature of the FPS sliding surface.

$$f_b=c_b\dot{x}_b+k_b x_b, \tag{8}$$

$$f_b=c_b\dot{x}_b+k_b x_b+(1-\alpha)F_0 W h_x, \tag{9}$$

$$f_b=k_b x_b+\mu_b W\,\mathrm{sgn}(\dot{x}_b). \tag{10}$$

where $\dot{x}_b$ = velocity of the base slab relative to the ground; h_x = nondimensional hysteretic displacement component [14]; and sgn(•) represents the signum function.

2.2. Numerical Assessment under Earthquake

Synthetic earthquakes compatible with a response spectrum of the Ethiopian standard, ES EN 1998-1:2015, are generated and used for non-linear response history analysis of the FB and BI building

models. Figure 3 shows the normalized elastic response spectra of the ES EN 1998-1:2015 for the five ground types specified in the standard. In this investigation, the design response spectrum corresponding to Ground Type E is considered to generate the synthetic earthquake time histories. Moreover, the ground acceleration corresponding to Seismic Zone IV has been implemented. The 1940 Imperial Valley (IV1940), 1994 Northridge (NR1994), and 1995 Kobe (KB1995) ground motion data (i.e., recorded at the El Centro, Sylmar, and Kobe Japan Meteorological Agency - KJMA Stations) are used as reference time histories to generate synthetic earthquakes. The peak ground accelerations (PGAs) of the considered earthquakes are 0.32 *g*, 0.60 *g*, and 0.83 *g*, respectively. Response spectrum compatible synthetic earthquakes are used in this study because of the absence of readily available recorded real earthquake data for the region. The considered synthetic earthquake data represent site-specific ground motions that reflect the expected seismic hazard accounting for local conditions.

The earthquake response of the six-story FB and BI building models are investigated under synthetic ground motions. The synthetic earthquakes are generated to be compatible with the response spectrum of the Ethiopian seismic standard for Ground Type E. The three synthetic earthquake time histories and their frequency content are presented in Figure 4, whereas the response spectra of the synthetic earthquakes are shown in Figure 5.

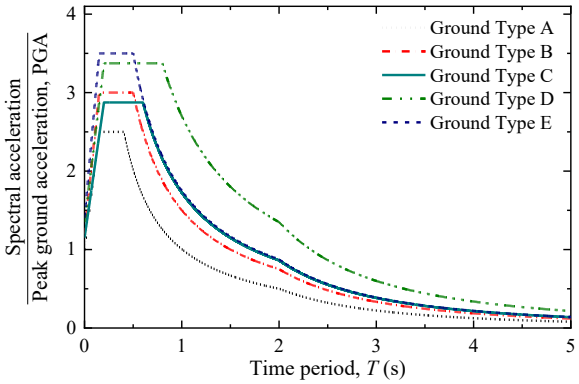


Figure 3 Normalized elastic response spectra of the ES EN 1998-1:2015 for the five ground types.

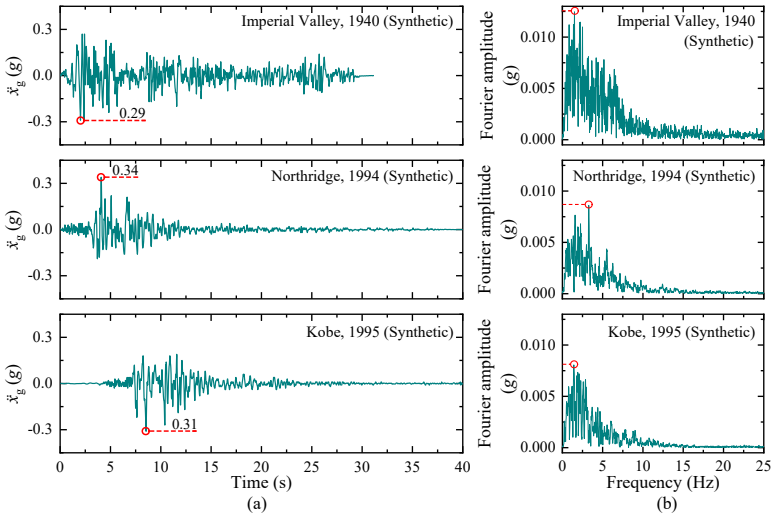


Figure 4 Details of the synthetic earthquakes compatible with the response spectrum of ES EN 1998-1:2015: (a) earthquake time histories and (b) frequency content.

Throughout the study, five buildings (i.e., two-, four-, six-, eight-, and ten-story buildings) with FB fundamental time period (T_s) values of 0.2 s, 0.4 s, 0.6 s, 0.8 s, and 1 s, respectively, are used. Also, the same buildings are isolated using elastomeric and sliding bearings, and evaluated under the three synthetic seismic ground motions. Here, T_b and ξ_b are considered to characterize the LRB. In contrast, the LCRB is characterized considering T_b, ξ_b, F_0, and q. Further, the

coefficient of friction (μ_b) and the isolator stiffness (k_b) are considered to characterize the FPS. The non-linear dynamic response history assessment of the multi-story buildings fitted with the LRB, LCRB, and FPS is performed using the state-space method. Four vibration response quantities of the multi-story buildings: (a) the top floor absolute acceleration, $\ddot{x}_{N,Abs}=\ddot{x}_N+\ddot{x}_b+\ddot{x}_g$; (b) normalized base shear, i.e., V_n, (c) inter-story drift ratio, and (d) isolator displacement, x_b, are studied. The values of V_n and the inter-story drift ratio of the j^{th} story (Δ_j) are computed as

$$V_n=V_1/W, \quad (11)$$

$$\Delta_j=(x_j-x_{j-1})/H_j \times 100, \quad (12)$$

where V_1 is the base shear; and H_j is the height of the j^{th} story, which is considered to be 3.5 m here.

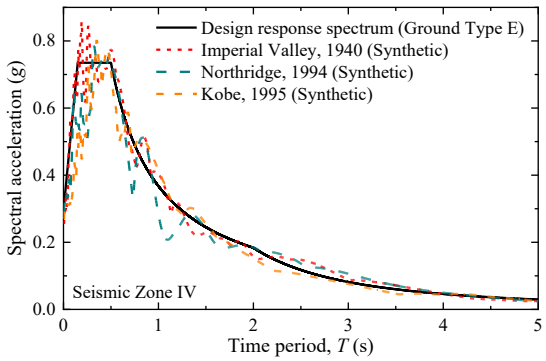


Figure 5 Response spectra of the synthetic earthquake time histories used in the investigation.

Table 1 outlines a summary of the key parameters of the superstructure and three types of isolators considered in the investigation.

Table 1 Key parameters of the superstructure and base isolators considered in the numerical assessment.

Structural Component		Parameter	Value(s) of the Parameter
Superstructure		Fixed-base fundamental time period, T_s^* (s)	0.2, 0.4, 0.6, and 0.8
		Superstructure damping ratio, ξ_s	0.02
		Story height, H_j (m)	3.5
Base isolator	LRB	Time period of isolation, T_b (s)	2, 2.5, 3, and 3.5
		Isolation damping ratio, ξ_b	2.5 % to 20 %
	LCRB	Time period of isolation, T_b (s)	2, 2.5, 3, and 3.5
		Isolation damping ratio, ξ_b	5 %
		Normalized yield strength, F_0	0.025 to 0.2
		Isolator yield displacement, q (cm)	2.5
	FPS	Time period of isolation, T_b (s)	2, 2.5, 3, and 3.5
		Friction coefficient of isolator, μ_b	0.025 to 0.2

The fundamental natural frequencies of the two-, four-, six-, eight-, and ten-story fixed-base buildings are 31.416 rad/s, 15.708 rad/s, 10.472 rad/s, 7.854 rad/s, and 6.283 rad/s, respectively. In contrast the fundamental natural frequencies of the five buildings isolated using LRB (T_b = 2.5 s) are 2.508 rad/s, 2.490 rad/s, 2.460 rad/s, 2.417 rad/s, and 2.364 rad/s, respectively. This shows that the introduction of base isolation has significantly reduced the natural frequencies of the buildings, especially for the short buildings.

The numerical study is conducted by performing two-dimensional (2D) non-linear time history analyses wherein the contributions of all modes of vibration are accounted for. Here, the structures are considered regular in plan and elevation. Also, the vertical vibration and torsional response of the structures due to the seismic action are considered to be negligible. Furthermore, the properties of

the superstructure and isolation systems are considered deterministic, and degradations in the properties of the structures and isolation systems are not considered. Therefore, the findings, discussions, and conclusions presented in the subsequent sections shall be viewed in consideration of the aforementioned assumptions.

3. RESULTS AND DISCUSSION

The merits of the base isolation method in the context of Ethiopian seismic standard, the influence of isolator properties on the vibration response, and the provisions of the seismic standard on base isolation are studied. The results are discussed in this section.

3.1. Efficacy of Base Isolation

The efficacy of the seismic isolation strategy in protecting structures is assessed considering the six-story multi-story building (T_s = 0.6 s, and ξ_s = 0.02), and the results are presented herein. The values of $\ddot{x}_{N,Abs}$, V_n, and x_b of the six-story BI buildings installed with the LRB, LCRB, and FPS are evaluated here. The time histories of the four response quantities of the six-story BI building models subjected to the synthetic Northridge, 1994 ground motion data are presented in Figure 6. The vibration response quantities of the FB building are depicted in the figure for comparison. The merit of isolating the superstructure from the ground motion using various base isolators is highlighted through the results shown in Figure 6. The presented results demonstrate that the placement of the isolators substantially reduces the seismic response of the buildings, i.e., $\ddot{x}_{N,Abs}$ and V_n, throughout the duration of the earthquakes.

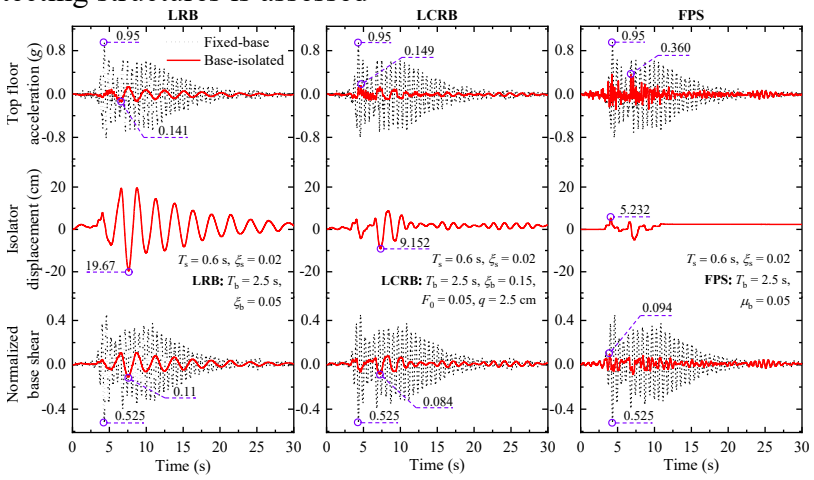


Figure 6 Time histories of floor acceleration, isolator displacement, and normalized base shear of the six-story BI building under the synthetic NR1994 earthquake.

The floor acceleration, $\ddot{x}_{N,Abs,p}$, of the FB building under the synthetic Northridge earthquake is obtained to be 0.95 *g*. It is seen in Figure 6 that the peak floor absolute acceleration is reduced to 0.141 *g*, 0.149 *g*, and 0.36 *g* for the buildings equipped with the LRB, LCRB, and FPS, respectively. The peak normalized base shear ($V_{n,p}$) is also reduced from 0.525 (for the FB building) to 0.11, 0.084, and 0.094 for the buildings installed with LRB, LCRB, and FPS, respectively. Table 2 presents a comparison of the peak values of $\ddot{x}_{N,Abs,p}$, $V_{n,p}$, inter-story drift ratio ($\Delta_{r,p}$), and isolator displacement ($x_{b,p}$) of the FB and BI buildings. It is observed that $\ddot{x}_{N,Abs,p}$, $V_{n,p}$, and $\Delta_{r,p}$ of the building are reduced significantly when seismic isolation is used. Based on the data presented in Table 2 on the building considered in this study, the average reductions in the $\ddot{x}_{N,Abs,p}$, $V_{n,p}$, and $\Delta_{r,p}$ are about 78 %, 83 %, and 88 %, respectively. Although the amount of reduction varies based on the type of isolator, earthquake, and isolator parameters, the reduction of

the seismic response quantities due to the base isolation is typically considerable.

Table 2 Peak response quantities of the six-story building with and without base isolation under the synthetic IV1940, NR1994, and KB1995 earthquakes (T_s = 0.6 s, T_b = 2.5 s).

Base Isolator	Ground Motion	Response Quantity	Fixed-Base	Base-Isolated	Response Reduction (%)
LRB T_b = 2.5 s, ξ_b = 0.1	IV1940 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	1.0582	0.1567	85.2
		$V_{n,p}$	0.5637	0.1075	80.9
		$\Delta_{r,p}$	0.0059	0.0007	88.4
		$x_{b,p}$ (cm)	-	18.3514	-
	NR1994 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	0.9508	0.1273	86.6
		$V_{n,p}$	0.5246	0.0987	81.2
		$\Delta_{r,p}$	0.0055	0.0006	88.7
		$x_{b,p}$ (cm)	-	17.0477	-
	KB1995 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	0.8624	0.1056	87.8
		$V_{n,p}$	0.5061	0.0806	84.1
		$\Delta_{r,p}$	0.0052	0.0005	90.4
		$x_{b,p}$ (cm)	-	13.7971	-
LCRB T_b = 2.5 s ξ_b = 0.05, F_0 = 0.05, q = 2.5 cm	IV1940 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	1.0582	0.1454	86.3
		$V_{n,p}$	0.5637	0.0903	84.0
		$\Delta_{r,p}$	0.0059	0.0006	89.3
		$x_{b,p}$ (cm)	-	9.5805	-
	NR1994 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	0.9508	0.1493	84.3
		$V_{n,p}$	0.5246	0.0836	84.1
		$\Delta_{r,p}$	0.0055	0.0006	89.6
		$x_{b,p}$ (cm)	-	9.1524	-
	KB1995 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	0.8624	0.1466	83.0
		$V_{n,p}$	0.5061	0.0920	81.8
		$\Delta_{r,p}$	0.0052	0.0006	87.8
		$x_{b,p}$ (cm)	-	10.5245	-
FPS T_b = 2.5 s, μ_b = 0.05	IV1940 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	1.0582	0.3428	67.6
		$V_{n,p}$	0.5637	0.1020	81.9
		$\Delta_{r,p}$	0.0059	0.0008	85.7
		$x_{b,p}$ (cm)	-	6.7799	-
	NR1994 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	0.9508	0.3596	62.2
		$V_{n,p}$	0.5246	0.0945	82.0
		$\Delta_{r,p}$	0.0055	0.0009	84.3
		$x_{b,p}$ (cm)	-	5.2317	-
	KB1995 (Synthetic)	$\ddot{x}_{N,Abs,p}$ (g)	0.8624	0.3744	56.6
		$V_{n,p}$	0.5061	0.0797	84.3
		$\Delta_{r,p}$	0.0052	0.0008	85.5
		$x_{b,p}$ (cm)	-	5.1688	-

The reductions in the vibration response quantities of the building are achieved because of the change in the dynamic behavior (shift in the fundamental frequency) of the building caused by the isolation systems. The increment in the isolation time period changes the natural frequency of the building away from the predominant frequencies of earthquakes, which reduces the earthquake energy transferred to the superstructure by avoiding resonance in those modes. The hysteretic behavior of the isolators also contributes to the dissipation of the energy imparted from the earthquakes. Accordingly, the findings presented in Figure 6 and Table 2 demonstrate that the seismic isolation technique considerably

reduces the seismic response of buildings considering the Ethiopian standard seismic provision. Therefore, the base isolation technique can suitably be used for earthquake protection of high importance structures, as per the Ethiopian standard, especially in regions of high seismicity.

The reduction in the lateral floor vibration as a result of seismic isolation improves the building behavior during earthquakes in various ways. Damage sustained by structural and non-structural components of BI buildings will be significantly reduced. This reduces the risk of loss of human life and injuries to a minimum, if any. In addition, it makes the BI buildings easily habitable after a large earthquake event and significantly reduces associated maintenance costs.

The technique also protects secondary systems such as expensive equipment and goods installed inside buildings from damage. This helps avoid unnecessary loss of money and assures continued functioning of the buildings and equipment during and after earthquakes, which is especially crucial for important structures, such as hospitals. The seismic protection benefits, as discussed herein, highlight the merits of considering base isolation in Ethiopia to protect important structures from earthquakes effectively. Therefore, the base isolation technique can be used as an effective alternative earthquake-resistant design technique for earthquake-prone areas of Ethiopia, especially for important and lifeline structures, such as hospitals, bridges, industrial structures, and power plants.

3.2. Effect of Base Isolation System Properties and Provisions of Ethiopian Standard

The influence of the properties of the three seismic isolators on the response quantities of the six-story BI building is investigated here. The influence of the isolator properties on the vibration response (i.e., $\ddot{x}_{N,Abs,p}$, $V_{n,p}$, $\Delta_{r,p}$, and $x_{b,p}$) is assessed through a detailed numerical investigation conducted by varying the isolation system characteristic parameters. The six-story building with T_s of 0.6 s and ξ_s of 0.02 is considered in the assessment. The isolation damping ratio of the LRB is varied from 2.5 % to 20 %, whereas F_0 values ranging between 0.025 and 0.2 are considered to model the LCRB. Moreover, the range of μ_b considered for the FPS is 0.025 to 0.2. The values of q and ξ_b of the LCRB are taken as 2.5 cm and 5 %, respectively.

The effect of the properties of the LRB, LCRB, and FPS on the $\ddot{x}_{N,Abs,p}$ of the BI building is presented in Figure 7. The results show that an increase in the ξ_b of the LRB typically causes the reduction of $\ddot{x}_{N,Abs,p}$. This indicates that, when a range of LRB damping ratios shall be considered during design, it is essential to consider the lower (minimum) possible value of the isolation damping ratio to compute the design floor acceleration response. Such a choice of the parameter is necessary to ensure the safety of the structure under the least favorable scenario. Justifiably, Section 10.8(2) of the seismic provisions of the Ethiopian standard, i.e., ES EN 1998-1:2015, specifies that the minimum value of the isolation damping shall be considered in the evaluation of acceleration response.

The influences of the F_0 of the LCRB and μ_b of the FPS on the $\ddot{x}_{N,Abs,p}$ are presented in Figure 7. For the LCRB, an increase in the value of F_0, in most cases, has caused in the increment of the $\ddot{x}_{N,Abs,p}$. However, it is also found that an increase in F_0 could, in some cases, result in the reduction of $\ddot{x}_{N,Abs,p}$. For instance, for the relatively small isolation time period value (i.e., T_b = 2 s) considered in this study, the $\ddot{x}_{N,Abs,p}$ is found to reduce with an increase in F_0 value up to about F_0 = 0.05. Beyond F_0 value of about 0.05, the floor acceleration response reverts to the increasing trend. For the FPS, it is seen that an increase in μ_b causes a consistent

increment of the top floor absolute acceleration response. Although Section 10.8(2) of ES EN 1998-1:2015 does not comment on the normalized yield strength of the LCRB, it is specified in the standard that the minimum value of μ_b shall be considered in the determination of the acceleration response of BI structures. Based on the findings presented in Figure 7, consideration of the minimum (smaller) value of μ_b leads to the consideration of a smaller value of acceleration in design. Therefore, choosing the minimum possible isolation friction coefficient does not account for the least favorable scenario and may lead to an unsafe design.

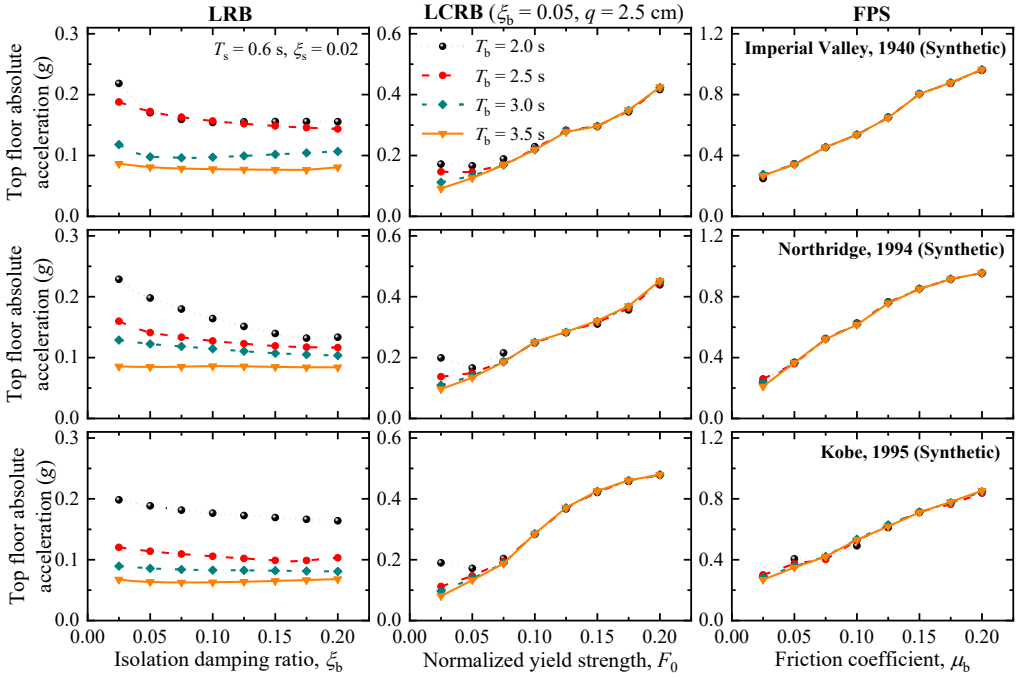


Figure 7 Effect of isolator parameters on the top floor acceleration of the six-story BI building under IV1940, NR1994, and KB1995 ground motions.

The influence of the properties of the three base isolators on the $V_{n,p}$ and $\Delta_{r,p}$ of the BI building is presented in Figure 8 and Figure 9, respectively. Both $V_{n,p}$ and $\Delta_{r,p}$ are found to reduce with an increase in the ξ_b of the LRB. This behavior necessitates the consideration of the minimum possible value of the isolation damping to quantify the design shear force in the building, which is consistent with the provision of Section 10.8(2) of ES EN 1998-1:2015. On the contrary, normalized base shear and inter-story drift reduce for an initial increment of F_0 of the LCRB (up to about F_0 = 0.05). Beyond F_0 = 0.05, the dynamic response quantities increase for an increase in the value of F_0.

In addition, it is established from the results that an increase in the FPS friction coefficient leads to an increment of $V_{n,p}$ and $\Delta_{r,p}$. The increasing trends of the dynamic response quantities under increasing values of F_0 and μ_b are attributed to the high-frequency vibration associated with large initial stiffness and large frictional resistance. However, the provisions of the ES EN 1998-1:2015 recommend using the minimum value of friction and do not seem to account for the undesirable effects of large frictional resistance fully. Here, it is important to note that results with similar implications have been reported in existing studies [20, 21]. Specifically, the findings of the optimization studies conducted by Jangid [20] and Rong [21] have shown that the optimum values of F_0 and μ_b suitable to achieve reduced floor accelerations are not

necessarily the larger values. This demonstrates that the findings of the current study are in alignment with those reported in previous studies.

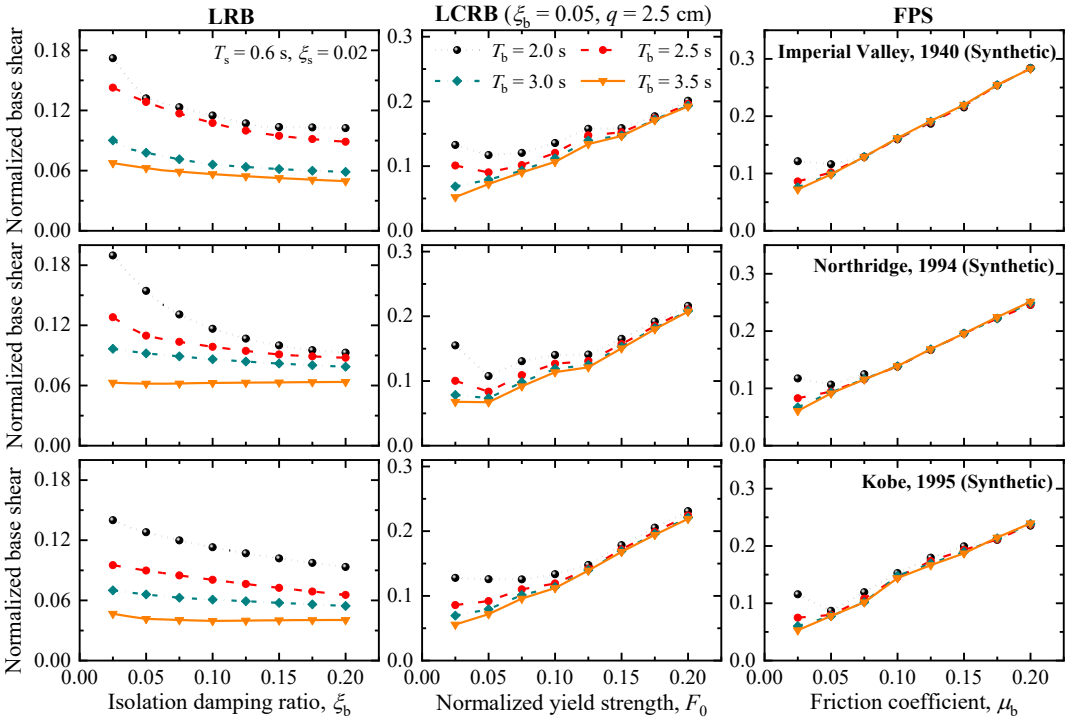


Figure 8 Effect of parameters of base isolators on the normalized base shear of the six-story BI building under IV1940, NR1994, and KB1995 ground motions.

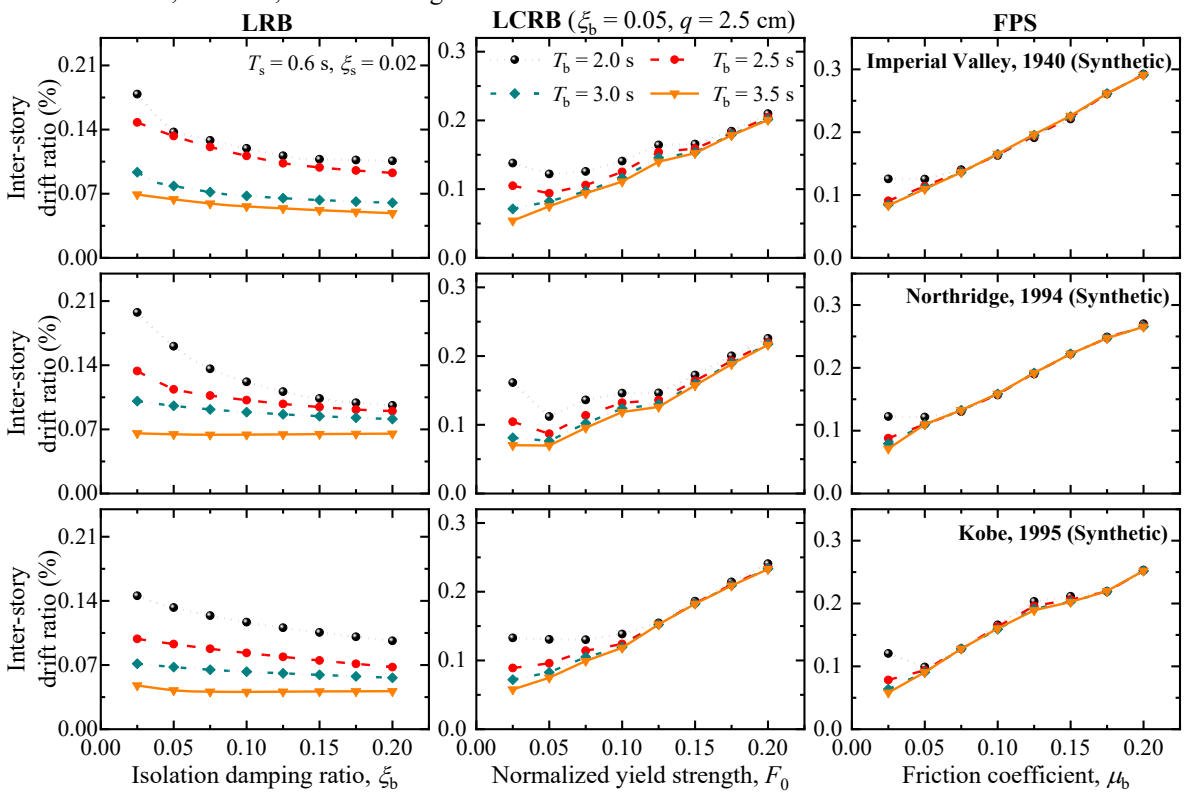


Figure 9 Effect of characteristic parameters of base isolators on the inter-story drift of the six-story BI building under IV1940, NR1994, and KB1995 ground motions.

The effect of the properties of the LRB, LCRB, and FPS on the isolator displacement of the six-story BI building is depicted in Figure 10. The results show that ξ_b of the LRB, F_0 of the LCRB, and μ_b of the FPS influence the isolator displacement response similarly. It is found that an increase in the three parameters results in the reduction of the isolator displacement. Accordingly, the critical design isolator displacement shall be computed considering the least favorable scenario of the minimum values of isolation damping ratio, normalized yield strength, and friction coefficient. The provision specified in Section 10.8(3) of ES EN 1998-1:2015 dictates that the displacement (isolator displacement response) should be evaluated considering the minimum values of isolation damping and coefficient of friction, which agrees with the findings presented here.

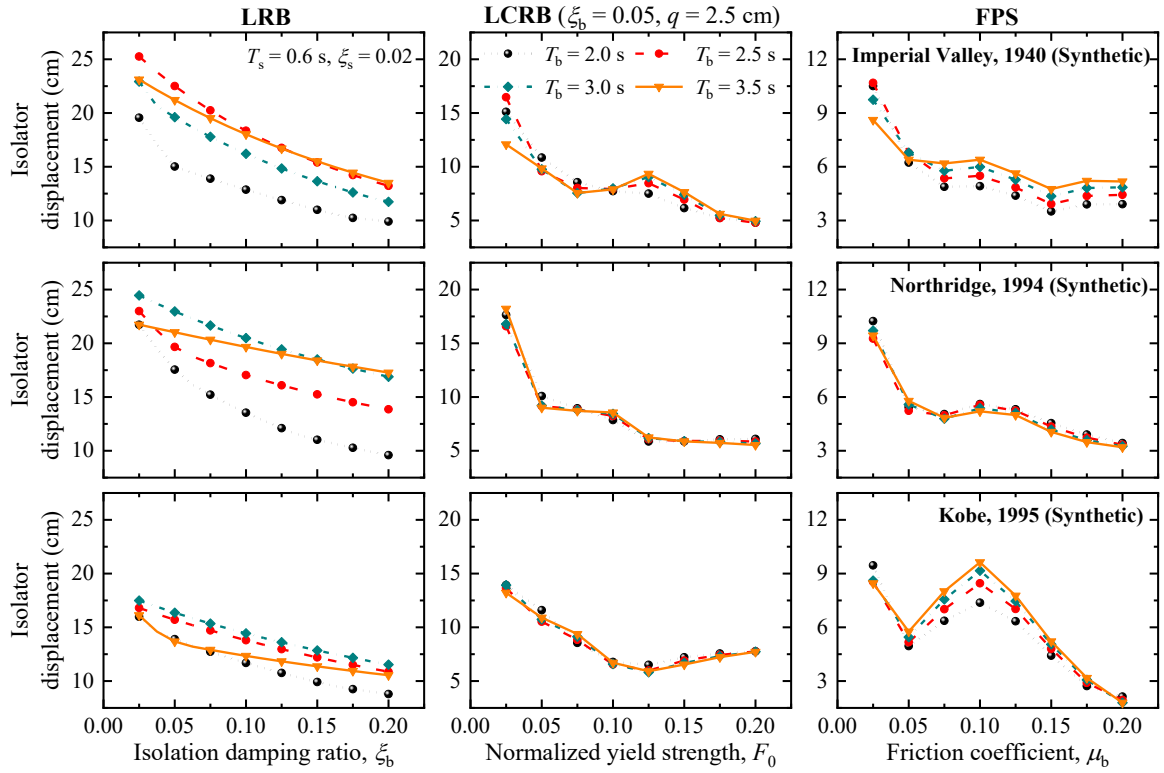


Figure 10 Effect of parameters of base isolators on the isolator displacement of the six-story BI building under IV1940, NR1994, and KB1995 ground motions.

3.3. Influence of Building Height on Choice of Isolator Parameter for Response Evaluation

The height of the building is one of the important factors that could influence the behavior of BI buildings. Therefore, it is essential to investigate its effect on the choice of the values of the isolator mechanical properties to be used for response evaluation. Five BI buildings with a different number of stories (N) are considered here to investigate the effect of building height. The FB fundamental time periods of the five buildings considered here are 0.2 s, 0.4 s, 0.6 s, 0.8 s, and 1 s for the two-, four-, six-, eight-, and ten-story buildings, respectively. Moreover, $\xi_s = 0.02$ is used for the five buildings, whereas $T_b = 2.5$ s is considered for all the BI buildings equipped with the LRB, LCRB, and FPS. In addition, ξ_b and q of the LCRB are taken as 0.05 and 2.5 cm, respectively.

Considering that each mechanical property of a base isolator could vary during the lifetime of a BI building, the quantification of the vibration response quantities of the BI building requires to be done

considering the values of isolator mechanical properties that result in the least favorable dynamic response. Each mechanical property of the base isolators with a mean value of χ_b could vary between a minimum value, i.e., $\chi_{b,Min}$, and a maximum value, i.e., $\chi_{b,Max}$. In practice, the quantification of a response quantity to be used for the design of the BI building shall be evaluated considering the least favorable value of the isolator property, either $\chi_{b,Min}$ or $\chi_{b,Max}$, whichever results in the worst scenario. Here, the response assessment of the five buildings equipped with LRB, LCRB, and FPS is conducted considering the minimum and maximum possible values of the respective mechanical properties of the isolators. For the LRB, the mean value of the isolation damping ratio is taken as $\xi_b = 0.1$. Accordingly, the minimum and maximum values of ξ_b are taken as $\xi_{b,Min} = 0.085$ and $\xi_{b,Max} = 0.115$, respectively, considering 15 % variation. For the LCRB, the minimum and maximum values of the normalized yield strength, i.e., $F_{0,Min}$ and $F_{0,Max}$, are taken as 0.0425 and 0.0575, respectively. Similarly, the minimum and maximum values of μ_b of the FPS, i.e., $\mu_{b,Min}$ and $\mu_{b,Max}$, are taken as 0.0425 and 0.0575, respectively.

The effect of the choice of the isolator property value, either $\chi_{b,Min}$ or $\chi_{b,Max}$, on four dynamic response quantities of BI buildings is investigated considering the five buildings having different heights, and the findings are presented in Figure 11. The four response quantities evaluated here are the $\ddot{x}_{N,Abs,p}$, $V_{n,p}$, $\Delta_{r,p}$, and $x_{b,p}$. The results depicted in Figure 11 show that the consideration of the minimum value of the isolation damping ratio of the LRB ($\xi_{b,Min}$) leads to larger values of the four response quantities than that of the case where the maximum isolation damping ratio ($\xi_{b,Max}$) is considered. Importantly, the results demonstrate that $\xi_{b,Min}$ causes in the unfavorable values of the four response quantities for all five buildings of different heights.

For the LCRB-controlled buildings, it is found that the unfavorable values of the peak top floor absolute acceleration, peak normalized base shear, and peak inter-story drift ratio are obtained when the maximum value of the normalized yield strength ($F_{0,Max}$) is considered. On the contrary, the minimum value of the normalized yield strength ($F_{0,Min}$) has resulted in the unfavorable value of the bearing displacement response for all the five buildings of different heights. For the buildings equipped with the FPS, the unfavorable values of the peak top floor absolute acceleration, peak normalized base shear, and peak inter-story drift ratio are obtained when the maximum value of the friction coefficient ($\mu_{b,Max}$) is considered. In contrast, the unfavorable value of the isolator displacement is associated with the consideration of the minimum value of the coefficient of friction of the FPS ($\mu_{b,Min}$) for all five buildings of different heights.

Yet again, the findings of the assessment conducted on the BI buildings of different heights (i.e., two-, four-, six-, eight-, and ten-story BI buildings) highlight that the isolator displacement response should be computed considering the minimum values of isolation damping ratio and coefficient of friction. These findings agree with the provision specified in Section 10.8(3) of ES EN 1998-1:2015. On the contrary, for all BI buildings of different heights considered in this study, the critical (unfavorable) values of the floor acceleration, base shear, and inter-story drift shall be obtained using the maximum values of the isolation damping ratio and coefficient of friction. This observation contradicts the provision specified in Section 10.8(2) of ES EN 1998-1:2015. In summary, it is established through the findings that the provision of

Section 10.8(2) of ES EN 1998-1:2015 does not necessarily agree with the expected behavior of BI buildings. Therefore, the provision must be carefully considered during the design of BI buildings. Finally, it is to be noted that provisions of other international seismic codes, such as AIJ-2016 [22] and ASCE 7-16 [23], dictate that the design of seismically isolated structures must consider the likely variations in the isolation device parameters. Particularly, the ASCE 7-16 [23] specifies that both the upper bound and lower bound properties of the isolation system shall be independently considered in the structural analyses of BI structures, the results of which are used to determine the governing demand parameters. These particular provisions are consistent with the findings of the current study and further reiterate the need for careful consideration of Section 10.8(2) of ES EN 1998-1:2015.

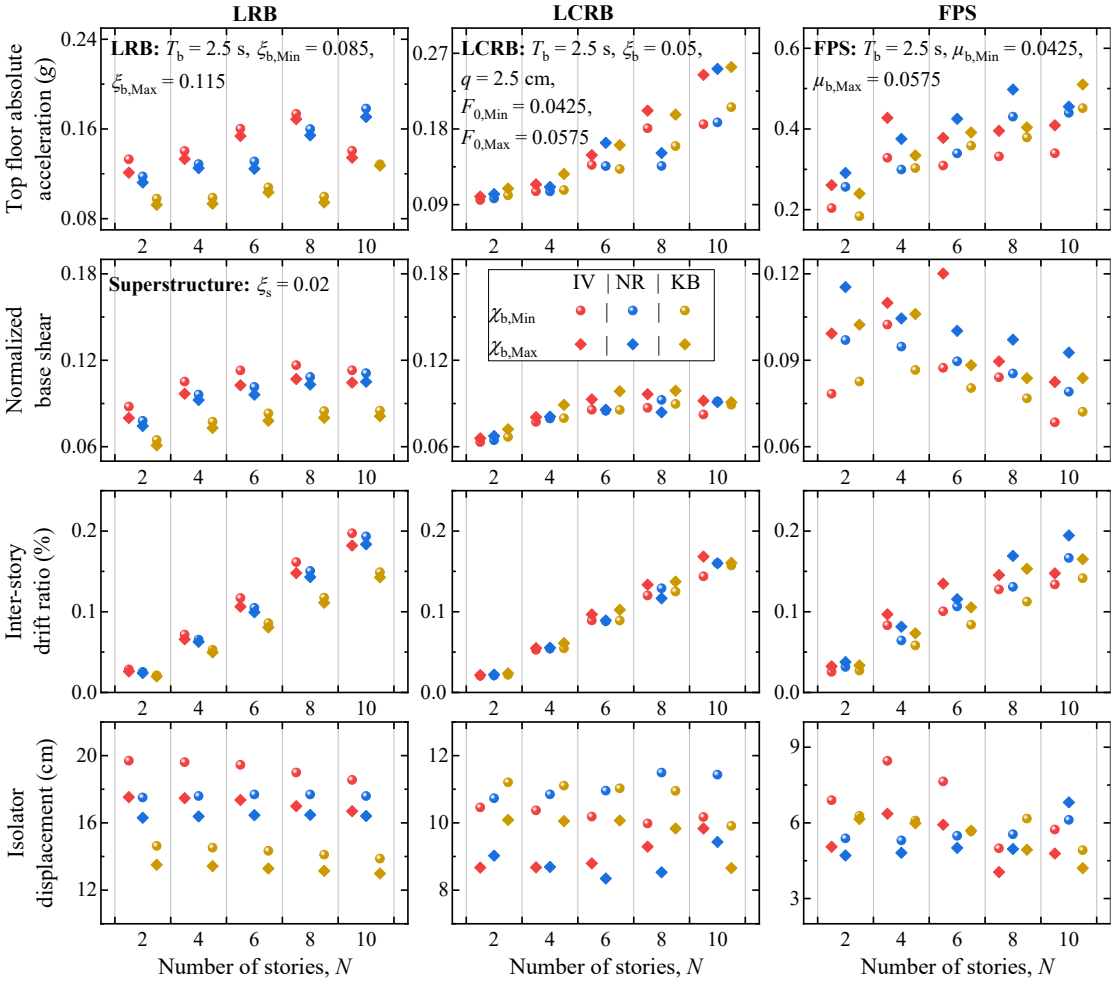


Figure 11 Comparison of the peak response quantities of BI buildings considering the minimum and maximum values of the mechanical properties of base isolators (i.e., LRB, LCRB, and FPS).

4. CONCLUSIONS

The efficacy of the base isolation technique in earthquake protection of multi-story buildings is studied considering the seismic provisions of the Ethiopian standard. The influence of isolator properties is investigated, and the findings are considered to evaluate the provisions of the Ethiopian seismic standard on consideration of base isolation system parameters. The following conclusions are drawn based on the findings.

Base isolation significantly reduces the seismic response quantities (i.e., the top floor absolute acceleration, base shear, and inter-story drift ratio) of the multi-story buildings.

For the six-story building considered in this study, the average reductions in the top floor absolute acceleration, normalized base shear, and inter-story drift achieved due to the implementation of base isolation are about 78 %, 83 %, and 88 %, respectively.

An increase in the damping ratio of the LRB causes the reduction of the peak top floor absolute acceleration, normalized base shear, and inter-story drift ratio. This behavior is consistent with the provisions of Section 10.8(2) of the seismic provisions of the Ethiopian standard (ES EN 1998-1:2015) in that the minimum value of the isolation damping shall be taken into account in the evaluation of acceleration response and shear forces in the structure.

For the building equipped with the LCRB, an increase in F_0 (yield strength) results in an initial reduction of the top floor absolute acceleration, normalized base shear, and inter-story drift ratio up to about $F_0 = 0.05$. Further increase in F_0 results in a consistent increment of the three response quantities.

An increase in the coefficient of friction of the FPS causes an increment of the top floor absolute acceleration, normalized base shear, and inter-story drift ratio response of the base-isolated building. Accordingly, to account for the least favorable scenario, the maximum possible value of the friction coefficient of the isolator shall be used in computing the design values of the three response quantities. On the contrary, Section 10.8(2) of ES EN 1998-1:2015 recommends the use of the minimum friction coefficient. Such a recommendation does not necessarily account for the undesirable effects of the frictional force and may lead to an inadequate design. Therefore, it is recommended that Section 10.8(2) of ES EN 1998-1:2015 be updated to consider the maximum friction coefficient as a possible scenario that likely causes the critical (design) values of the three response quantities.

Increase in the damping ratio of the LRB, normalized yield strength of the LCRB, and friction coefficient of the FPS results in the reduction of the isolator displacement. Therefore, the evaluation of the critical design isolator displacement shall be computed considering the least favorable scenario of the minimum values of isolation damping ratio, normalized yield strength, and friction coefficient, which is consistent with the recommendation of Section 10.8(3) of ES EN 1998-1:2015.

The LRB, LCRB, and FPS can be used as effective isolation systems for the protection of important structures in earthquake-prone areas of Ethiopia. However, the provisions of Section 10.8(2) of ES EN 1998-1:2015 must be carefully considered during design.

In summary, the base isolation technique, which delivers superior seismic protection of important structures, is recommended as an excellent strategy to help achieve enhanced community resilience against earthquakes. Notable focus areas for future research in the domain include the economic feasibility of base isolation and the development of comprehensive guidelines for testing and monitoring the behavior of isolation devices in the Ethiopian context.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support received by Daniel Habtamu Zelleke from Haramaya University and the

Ministry of Education, Ethiopia, during the execution of this research work.

REFERENCES

[1] K. K. Alaneme, and E. A. Okotete, "Critical Evaluation of Seismic Activities in Africa and Curtailment Policies - A Review," *Geoenvironmental Disasters*, vol. 5, no. 1, 2018.

[2] A. Ayele, K. Woldearegay, and M. Meten, "A Review on the Multi-Criteria Seismic Hazard Analysis of Ethiopia: With Implications of Infrastructural Development," *Geoenvironmental Disasters*, vol. 8, no. 9, pp. 43-55, 2021.

[3] A. Worku, "Recent Developments in the Definition of Design Earthquake Ground Motions Calling for a Revision of the Current Ethiopian Seismic Code - EBCS 8: 1995," *Journal of Ethiopian Engineers and Architects*, vol. 28, no. 1, pp. 1-15, 2011.

[4] S. Kinde, "Proposed Considerations for Revision of EBCS-8:1995 for Conservative Seismic Zoning and Stringent Requirements for Torsionally Irregular Buildings," *Journal of Ethiopian Engineers and Architects*, vol. 20, pp. 15-22, 2003.

[5] S. Kinde, S. Engeda, A. Kebede, and E. Tessema, "Notes and Proposed Guidelines on Updated Seismic Codes in Ethiopia: Implication for Large Scale Infrastructure," *Journal of Ethiopian Engineers and Architects*, vol. 28, no. 1, pp. 91-110, 2011.

[6] EBCS 8, "Design of Structures for Earthquake Resistance," *Ethiopian Building Code Standard* (*EBCS*), Ministry of Works and Urban Development, Addis Ababa, Ethiopia, 1995.

[7] USGS, "Earthquake Lists, Maps, and Statistics," *United States Geological Survey* (*USGS*), 2025. https://www.usgs.gov/natural-hazards/earthquake-hazards/lists-maps-and-statistics [Accessed May 10, 2025].

[8] ES EN 1998-1:2015, "Design of Structures for Earthquake Resistance - Part 1: General Rules - Seismic Actions and Rules for Buildings," *Ethiopian Standards*, Ministry of Construction, Ethiopia, 2015.

[9] J. M. Kelly, "Aseismic Base Isolation: Review and Bibliography," *Soil Dynamics and Earthquake Engineering*, vol. 5, no. 3, pp. 202-216, 1986.

[10] J. M. Kelly, and S. B. Hodder, "Experimental Study of Lead and Elastomeric Dampers for Base Isolation Systems in Laminated Neoprene Bearings," *Bulletin of the New Zealand National Society for Earthquake Engineering*, vol. 15, no. 2, pp. 53-67, 1982.

[11] D. M. Fenz, and M. C. Constantinou, "Behaviour of the Double Concave Friction Pendulum Bearing," *Earthquake Engineering and Structural Dynamics*, vol. 35, pp. 1403-1424, 2006.

[12] S. P. Chang, N. Makris, A. S. Whittaker, and A. C. Thompson, "Experimental and Analytical Studies on the Performance of Hybrid Isolation Systems," *Earthquake Engineering and Structural Dynamics*, vol. 31, pp. 421-443, 2002.

[13] R. S. Jangid, "Optimum Lead-Rubber Isolation Bearings for Near-Fault Motions," *Engineering Structures*, vol. 29, pp. 2503-2513, 2007.

[14] D. H. Zelleke, S. Elias, V. A. Matsagar, and A. K. Jain, "Supplemental Dampers in Base-Isolated Buildings to Mitigate Large Isolator Displacement under Earthquake Excitations," *Bulletin of the New Zealand Society for Earthquake Engineering*, vol. 48, no. 2, pp. 100-117, 2015.

[15] N. Kani, M. Takayama, and A. Wada, "Performance of Seismically Isolated Buildings in Japan: Observed Records and Vibration Perception by People in Buildings with Seismic Isolation," *Proceedings of the* 8th *US National Conference on Earthquake Engineering*, San Francisco, California, United States of America (USA), Paper No.: 2181, 2006.

[16] M. Forni, A. Poggianti, and A. Dusi, "Seismic Isolation of Nuclear Power Plants," *Proceedings of the* 15th *World Conference on Earthquake Engineering*, Lisbon, Portugal, 2012.

[17] D. H. Zelleke, S. K. Saha, and V. A. Matsagar, "Reliability-Based Multi-Hazard Design Optimization of Base-Isolated Buildings," *Engineering Structures*, vol. 301, pp. 117242, 2024.

[18] G. C. Hart, and K. Wong, *Structural Dynamics for Structural Engineers*, John Wiley and Sons Inc., 605 Third Avenue, New York, 2000.

[19] Y. K. Wen, "Method for Random Vibrations of Hysteretic Systems," *Journal of the Engineering Mechanics Division*, vol. 102, no. 2, pp. 249-263, 1976.

[20] R. S. Jangid, "Optimum Frictional Elements in Sliding Isolation Systems," *Computers and Structures*, vol. 76, no. 5, pp. 651-661, 2000.

[21] Q. Rong, "Optimum Parameters of a Five-Story Building Supported by Lead-Rubber Bearings under Near-Fault Ground Motions," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 39, no. 1, pp. 98-113, 2020.

[22] AIJ-2016, "Design Recommendations for Seismically Isolated Buildings," *Architectural Institute of Japan* (*AIJ*), 2016.

[23] ASCE 7-16, "Minimum Design Loads and Associated Criteria for Buildings and Other Structures," *American Society of Civil Engineers* (*ASCE*), 2016.

Investigation of the Effect of Lime Treatment on the Soil Water Characteristics Curve of Expansive Soils

Muna Mohammed Seid[1] and Tezera Firew Azmatch[2,*]

[1] *Faculty of Civil and Water Resource Engineering, Bahirdar Institute of Technology, Bahiradar, Ethiopia*
[2] *School of Civil and Environmental Engineering, College of Technology and Built Environment Addis Ababa University, Addis Ababa, Ethiopia.*
[*]*Corresponding author's E-mail address: tezera.firew@aait.edu.et*

DOI: https://doi.org/10.63990/zede.v43i.12972

ABSTRACT

The engineering behavior of natural untreated and compacted lime-treated expansive soils, which typically exist in a state of unsaturated condition, can be better-explained using concepts from unsaturated soil mechanics. The soil water characteristics curve (SWCC) is the key unsaturated soil property for obtaining unsaturated soil property functions (USPFs). However, there are limited studies on the effect of lime on the SWCC of lime-treated expansive soils. This study investigated the effect of lime on the SWCC of lime-treated expansive soils. The drying portion of SWCCs for untreated natural soil and lime - treated soil samples with three different lime contents (3 %, 6 % and 9 %) with 7 days of curing were studied. The SWCCs were determined by using pressure plate apparatus in the suction range of 0 – 1400 kPa. The shrinkage curve (SC) was also determined to evaluate the change in volume of the different soil samples. The experimental results indicate that SWCC is affected by lime treatment and there is a change in the SWCC parameters and in the shape and position of SWCC as the percentage of lime is changed. The SWCC of the lime-treated soil samples show a higher rate of desaturation as the lime content increases. The Air Entry Value (AEV) and residual water content of lime-treated soil decreases with increase in percentage of lime and the SWCC shifts towards the left side as the AEV decreases. The differences in AEV obtained from gravimetric- water-content-based SWCC (w-SWCC) and degree-of-saturation-based SWCC (S-SWCC) for the lime-treated soil samples were small when compared to the untreated natural soil sample.

Keywords: Air-entry value, Expansive soil, Lime treatment, Shrinkage curve, Soil–water characteristic curve, Unsaturated soil,

1. INTRODUCTION

Expansive soils experience large volumetric changes when subjected to change in moisture content. Infrastructures, particularly light-weight structures are severely damaged by these volumetric changes. This problem of expansive soils can be addressed by applying chemical treatment, such as lime treatment for stabilization of the soil.

Researchers have investigated the swell-shrink response of expansive soils using index properties and other soil laboratory tests. However, the engineering behavior of compacted expansive soils that are typically in a state of unsaturated condition can be better-interpreted if the influence of matric suction is considered [1]. In order to establish the relationship between unsaturated soil theory and engineering problems related to expansive soils, which exist at unsaturated state, the use of SWCC of expansive soils is investigated. The SWCC constitutes the primary soil information required for the analysis of seepage, shear strength, volume change, air flow, and heat flow problems involving unsaturated soils. The SWCC is relatively easy to measure and has become the key unsaturated soil property for obtaining USPFs.

Following the development of unsaturated soil mechanics, many researches were carried out in measuring the SWCC of soils [2, 3, 4, and 5]. Most of these studies were mainly on either non-expansive soils or on untreated natural expansive soils. However, very limited experimental data have been reported in literature on the SWCC of chemically treated expansive soils [6], especially on the impact of lime on the SWCC of expansive soils and hence on the USPFs of expansive soils.

The effects of lime treatment on the physicochemical properties of expansive soils can be attributed to a number of chemical reactions altering the soil nature and structure. The main two chemical reactions are [6, 7] are: (i) an immediate ion exchange reaction between exchangeable clay ions and calcium ions provided by the lime. Flocculation and agglomeration of the soil particles occur, transforming the plastic soil to a more granular and less plastic material [6, 8]. As a result of this reaction, the soil generally acquires an aggregated, more porous and less deformable structure [6, 9, 10]. In the context of SWCC studies, this would be expected to affect water retention. (ii) long-term pozzolanic reactions which promote dissolution of siliceous and aluminous compounds from the clay mineral lattice, reacting with calcium ions in the pore water to form calcium silicate hydrates, calcium aluminate hydrates and hydrated calcium aluminosilicates. This can potentially change soil pore connectivity as well as pore size distribution [10, 11].A simplified qualitative view of typical soil-lime reactions [7] is as follows:

$$Ca(OH)_2 \rightarrow Ca^{2+} + 2(OH)^- \quad (1)$$

$$Ca^{2+} + 2(OH)^- + SiO_2(\text{clay silica}) \rightarrow C{-}S{-}H \quad (2)$$

$$Ca^{2+} + 2(OH)^- + Al_2O_3(\text{clay alumina}) \rightarrow C{-}A{-}H \quad (3)$$

Where, C = CaO, A = Al_2O_3, and H = H_2O

Although it is clear that lime treatment, as indicated above, can potentially affect the water retention of the lime-treated soil, research on the SWCC of lime-treated expansive soils is relatively limited [6, 12]. In this study, hydrated lime with different contents was added into expansive soil. The optimum water contents and maximum dry density of the soil samples were determined by standard compaction method. The drying SWCC tests were carried out by using pressure plate apparatus. The Shrinkage Curve (SC) were determined for both untreated and lime-treated soil samples using ring method to account for the volume change behavior of expansive soils upon drying. The SC was combined with gravimetric-water-content-based SWCC (w-SWCC) to determine the degree-of-saturation-based SWCC (S-SWCC). The effect of lime treatment on the SWCC parameters of expansive soil was investigated using the w-SWCC and S-SWCC.

2. MARERIALS AND METHODS

Soil samples were collected from a test pit located at Ayer- Tena area in Bahirdar, Ethiopia, from a depth of 1.5m. Soil samples were prepared according to [13]. Tests were conducted to determine the index properties of the soil samples using ASTM standards. The index properties of the natural untreated soil are presented in Table 1.

Table 1 Properties of expansive soil used in this study

Clay (%)	85.9
Sand (%)	3.8
Silt (%)	9.8
Liquid Limit (%)	108.1
Plastic Limit (%)	38.0
Plasticity Index (%)	70.1
Free Swell (%)	135
Specific gravity Gs	2.73
pH	7.5
USCS Soil Classification	CH

The chemical composition of hydrated lime used in this investigation was studied by using X-Ray Fluorescence analysis [14] and is presented in Table 2.

Table 2 Chemical composition of hydrated lime used in this study

Constituent	Percentage by weight (%)
SiO_2	6.21
Al_2O_3	2.18
Fe_2O_3	3.57
CaO	59.47
MgO	3.91
Na_2O	0.61
K_2O	0.79
TiO_2	0.33
P_2O_5	0.21
MnO	0.28
SO_3	0.58

The minimum necessary percentage of lime to treat this soil was determined using pH test as per the procedure outlined in [15]. According to this test, the lime percentage needed for soil stabilization was found to be 6 %. The lime requirement determined by a pH test (6 %), as well as additional two different lime content values below and above 6 % (i.e., 3 % and 9 %) were applied to evaluate the effect of lime treatment on the SWCC. After mixing thoroughly the dry powders of the soil and lime, water was carefully added in small increments, and the wet paste was mixed thoroughly. The soil samples were allowed to cure by covering them with impermeable plastic bag. Lime-treated soil samples were compacted with their respective Maximum Dry Density (MDD) and Optimal Moisture Content (OMC) values and covered with impermeable plastic bag and then cured for 7 days.

The SWCCs were determined for untreated natural soil and lime-treated soil samples with lime content of 3 %, 6 %, and 9 % for 7-day curing periods following the drying path as per [16] using pressure plate apparatus. The pressure plate apparatus is a reliable and widely used method of measuring SWCC [17]. The measurement capacity of the pressure plate apparatus is governed by the air-entry value of the ceramic disc which is typically limited to 1500 kPa [6]. This study has applied suction ranges of 0 – 1400 kPa to measure SWCC, since the air entry value of the available ceramic disc was 1500 kPa. The soil samples are initially saturated for the SWCC and SC tests. SC of untreated natural soil and lime-treated samples with 3 %, 6 %, and 9 % for 7-days of curing were measured with a digital micrometer.

3. RESULTS AND DISCUSSION

3.1 Moisture – Density Relationships

The soil samples were prepared by compacting in the Standard Proctor Test using a 2.5 kg hammer falling 308 mm into a soil-filled mold. The mold was filled with three equal layers of soil. Each soil-filled mold received 25 drops of hammer. The dry density vs. moisture content curve was then plotted to determine the MDD and OMC. Untreated natural and lime-treated soil samples yielded the typical bell-shaped compaction curves as shown in Figure 1. The MDD decreased and the OMC increased as percentage of lime increases. MDD for an untreated natural soil was 1.26 g/cm^3 and reduced to 1.24 g/cm^3, 1.21 g/cm^3 and 1.18 g/cm^3 with the addition of 3 %, 6 % and 9 % lime, respectively. The OMC for an untreated natural soil was 33.8 % and increased to 35.6 %, 40.3 % and 41.2 % with the addition of 3 %, 6 % and 9 % lime, respectively. This is due to the fact that hydrated lime is finer than soil, which increases surface area, as well as flocculation and agglomeration, which cements soil particles and creates greater pore structure [6]. Wetting the fine lime's large surface area necessitates the addition of extra water. As a result, the MDD is reduced while the OMC is increased.

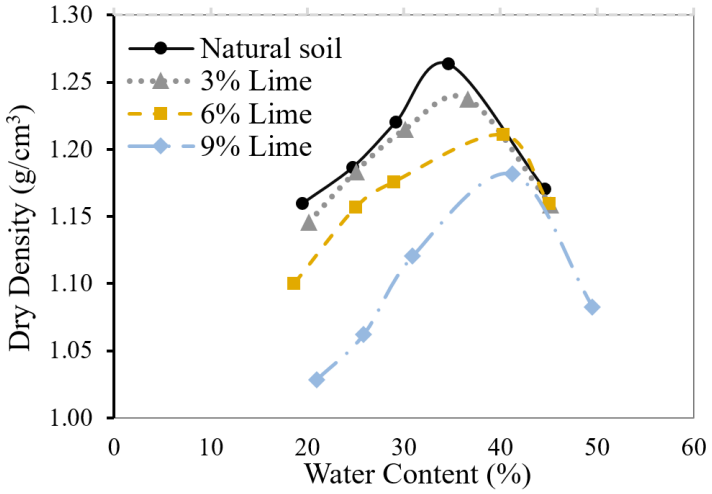


Figure 1 Compaction curves for untreated natural and lime-treated soil sample

3.2 SWCC and SC: Curve Fitting and Analysis

3.2.1 The fitting of SC

Even though several equations for fitting SCs have been suggested by different researchers, the hyperbolic equation [1], Eq. (4), provides the best match for the curves, while w-SWCC is combined with the SC test [18] and the fitting parameters are obtained using the EXCEL Solver function.

$$e(\omega) = a_{sh}\left(\left(\frac{\omega}{b_{sh}}\right)^{c_{sh}} + 1\right)^{\frac{1}{c_{sh}}} \qquad (4)$$

where a_{sh} = minimum void ratio upon complete drying (i.e., ranging from 0.4 to 1.0), b_{sh} = variable related to the slope of the drying curve calculated as: $b_{sh} = (a_{sh} \times S_o)/G_s$
and c_{sh} = variable related to the sharpness of curvature as the soil desaturates, and S_o = initial degree of saturation.

3.2.2 The fitting of w-SWCC and S-SWCC

In order to analyze the effect of lime addition on the SWCC parameters of expansive soils, the model by Fredlund and Xing [19] was used to fit the curves. The fitting formula for w-SWCC is expressed in Eq. (5) and the fitting parameters were obtained using the EXCEL Solver function [8].

$$\omega(\psi) = c(\psi)\frac{\omega_s}{\left\{\ln\left[e + (\psi/a)^n\right]\right\}^m} \qquad (5)$$

Where:

$\omega(\psi)$= gravimetric water content at any soil suction, ψ, e = constant equal to 2.71828, ω_s = saturated gravimetric water content, a = fitting parameter indicating the inflection point that bears a relationship to the air-entry value, n= fitting parameter related to the rate of desaturation, m = fitting parameter related to the curvature near residual conditions and $C(\psi)$= correction factor directing the SWCC to 10^6 kPa at zero water content, and given by Eq. (6):

$$C(\psi) = 1 - \frac{\ln(1+\psi/\psi_r)}{\ln\left[1+(10^6/\psi_r)\right]} \qquad (6)$$

The w - SWCC paired with SC was used to determine S-SWCC, which was employed to identify the true AEV. The void ratio and water content provided by the SC can be used to calculate the degree of saturation, which can then be used to convert the w-SWCC to the S-SWCC [20]. The fitting formula for S-SWCC is expressed as Eq. (7) and the fitting

parameters were obtained using the EXCEL Solver function.

$$S(\psi) = c(\psi)\frac{S_o}{\left\{\ln\left[e+\left(\psi/a\right)^n\right]\right\}^m} \quad (7)$$

Where S_o is the initial degree of saturation, $S(\psi)$ is degree of saturation at any suction calculated as: $S(\psi) = \frac{G_s\omega(\psi)}{e(\omega(\psi))}$, with G_s is specific gravity, $\omega(\psi)$ is measured gravimetric water content determined using pressure plate test and $e(\omega(\psi))$ is calculated void ratio using Equation 1 and the other parameters are described in Eq. (5).

3.3 Analysis of Soil Water Characteristic Curves

3.3.1 Analysis of w-SWCCs

Curing periods of 7-day and 14-day were initially investigated. The results indicated that the SWCCs for 7-day curing and 14-day curing are almost the same as presented in Table 3. Hence, further investigation was carried out using 7-day curing period. The effect of lime addition on w-SWCC of soil samples cured for 7-days is presented in Table 3 and Figure 2. The impact of lime addition on expansive soils was investigated by observing changes to Fredlund and Xing [19] SWCC curve fitting parameters. The results indicate that the SWCC curve fitting parameters were affected by the addition of lime. The addition of lime changed the shape and position of the w-SWCCs significantly. The value of "a" and AEV increased at first, but as the percentage of lime increased, the value of "a" and AEV decreased. The AEV first increased from the natural-untreated value of 75 kPa to 130 kPa at a lime content of 3 %, it then decreased to 115 kPa and 80 kPa at lime contents of 6 % and 9 %, respectively. The "n" value of a lime-treated soil sample first decreased, but then began to increase as the percentage of lime content increased.

Table 3 w-SWCC fitting parameters of Fredlund and Xing model [19]

Parameters	Natural	7-day curing period			14-day curing period		
		3%	6%	9%	3%	6%	9%
ω_s (%)	45.36	42.36	41.02	40.56	42.29	40.96	40.51
a	133.75	295.88	235.37	115.98	295.88	235.37	115.98
n	1.78	1.49	1.72	2.56	1.49	1.72	2.56
m	0.31	0.60	0.51	0.40	0.60	0.51	0.40
ψ_r (kPa)	1780	1449	1449	1402	1449	1449	1402
R^2	0.999	0.999	0.999	0.999	0.999	0.999	0.999
AEV(kPa)	75	130	115	80	130	115	80
ω_r (%)	25.08	21.32	20.11	17.33	21.25	20.05	17.28

The w-SWCC of the lime-treated soil samples shows a higher rate of desaturation as the lime content increases, compared to the untreated natural soil sample, implying that the water from the pores was easily drained in lime-treated soils when suction increased, compared to the untreated soil sample. The "m" value of a lime-treated soil sample first increased, but then decreased as the proportion of lime content increased. Both saturated and residual water contents decreased as the lime content increased.

3.3.2 Analysis of SCs

The effect of lime addition on SC of natural soil and lime-treated soil samples cured for 7 days is presented in Figure 3 and Table 4. The results indicate that the

curve fitting parameters are affected by lime treatment.

When comparing the SCs of the lime-treated soil sample to the SCs of the untreated natural soil sample, it was noted that the SCs of the lime-treated soil sample shifted upward. This is due to the addition of lime, which resulted in a larger void ratio in the lime-treated soil sample. The "a_{sh}", "b_{sh}", and "c_{sh}" value increased up to 7-day curing period for all lime-treated soil samples at the percentage lime considered in this study. For lime-treated soil samples, "c_{sh}" value was higher and considered as low compressible soils, which essentially produced a horizontal line from the initial void ratio to completely dry void ratio. On the other hand, untreated natural soil sample produced a gradual curve that immediately started to curve from the completely dry void ratio and gradually moved towards saturation line, which would have much higher compressibility (or undergo considerable volume change) up-on drying.

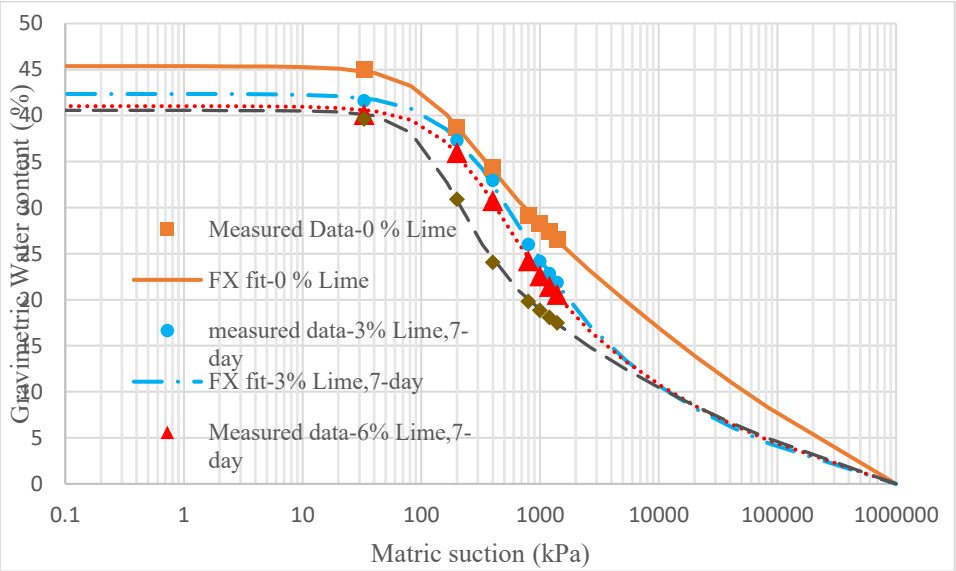


Figure 2 w-SWCC for untreated natural soil sample and lime-treated soil sample with 7-day of curing period.

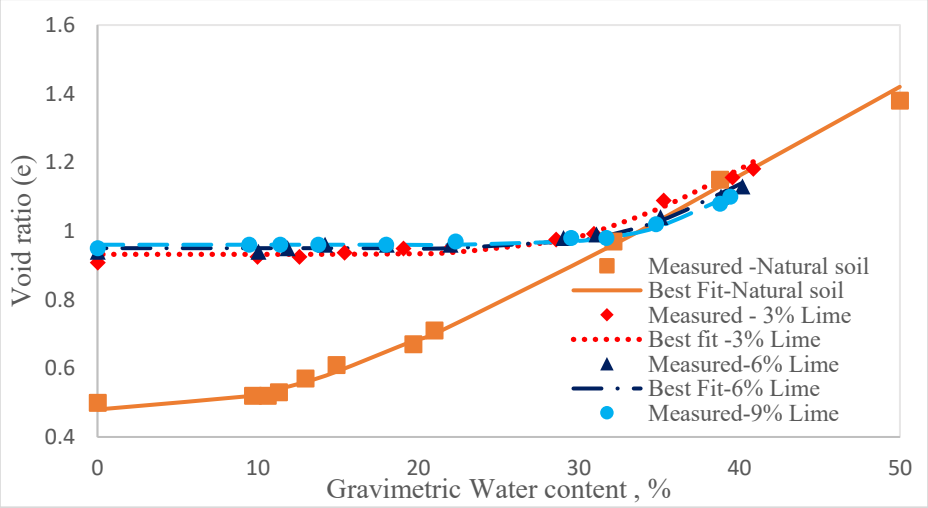


Figure 3 SC for untreated natural soil sample and lime- treated soil sample with 7-day of curing

Table 4 SC fitting parameters for untreated and lime-treated soil samples

Parameters	Natural	7-day curing period		
		3%	6%	9%
ω_s (%)	49.99	41.63	40.17	39.43
a_{sh}	0.48	0.93	0.95	0.96
b_{sh}	0.17	0.32	0.34	0.34
C_{sh}	2.56	8.04	10.74	14.69
R^2	0.997	0.980	0.983	0.985

3.3.3 Analysis of S-SWCCs

For soils that do not experience considerable volume change when soil suction increases, the AEV determined using w-SWCC and S-SWCC is the same. However, for a soil that undergoes significant volume change as soil suction increases, the AEV depends on the w-SWCC and SC test results. In such soils, the AEV is underestimated by the w-SWCC test result. In such circumstances, SWCC paired with SC should be employed to identify the true AEV [18]. Table 5 and Figure 4 present the S-SWCC results for the soil samples.

Table 5 S-SWCC fitting parameters of Fredlund and Xing model [19]

Parameters	**Natural**	**7-day curing period**		
		3%	**6%**	**9%**
S_o (%)	95.70	89.19	87.66	86.63
a	1200.00	524.85	430.22	213.59
n	0.18	3.69	4.20	4.63
m	0.17	0.19	0.20	0.22
ψ_r (kPa)	3358	1598	1213	1100
R^2	0.994	0.999	0.999	0.999
AEV (kPa)	1200	245	210	150
S_r (%)	79.61	60.80	58.46	49.46

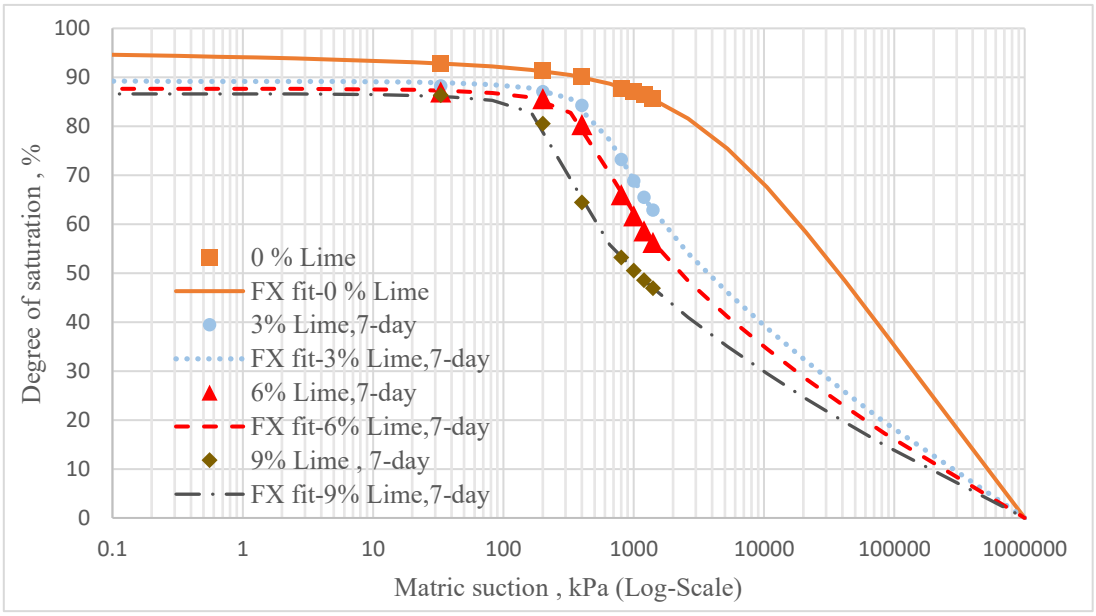


Figure 4 S-SWCC for untreated natural soil sample and lime- treated soil sample with 7-day of curing period.

The shape and position of the S-SWCCs changed significantly as a result of the addition of lime, as shown in Figure 4. The S-SWCC of untreated natural soil has a smooth desaturation slope across the suction range, whereas the S-SWCC of lime-treated soils exhibits a steep desaturation slope between the AEV and residual suction. The impact of lime addition on expansive soils was investigated by observing changes to Fredlund and Xing [19] SWCC model parameters. The results indicate that the SWCC curve fitting parameters were affected by the addition of lime. The value of "a" decreased as percentage of lime increased. Similarly, the AEV decreased as percentage of lime increased. It decreased from 1200 kPa at natural-untreated value to 245 kPa, 210 kPa and 150 kPa, at lime contents of 3 %, 6 %, and 9%, respectively, for seven days of curing. A reduced AEV in lime-treated soil samples is related to a lower water retention at low suctions. This is a result of the immediate ion exchange reaction between exchangeable clay ions and calcium ions provided by the lime, which results in flocculation and agglomeration of the soil particles that transform the plastic soil to a more granular and a more porous microstructure [6]. This indicates that the lime treatment has produced a more open microstructure with larger void ratio and hence a lower AEV. The "n" value of a lime-treated soil sample increases as the percentage of lime increases, meaning that the addition of lime results in a more uniform pore size distribution [21] and a faster rate of desaturation than a natural soil sample that has not been treated. The effects of lime treatment on the values of "m" were found to be insignificant. The residual suction (ψ_r) value decreased as the percentage of lime increased. It decreased from 3368 kPa at natural-untreated value to 245 kPa, 210 kPa and 110 kPa, at lime contents of 3 %, 6 %, and 9 %, respectively. Both initial and residual degree of saturation decreased as the lime content increased. This is due to the fact that lime-treated soils absorb less water than untreated natural soils at saturation stage. However, as the suction increases, the rate of desaturation increases, and water from the pores is expelled faster due to more open structures generated by the addition of lime [6].

4. CONCLUSIONS

The results in this study indicate that SWCC was affected by lime treatment and there were changes in the SWCC parameters, the shape and position of SWCC, as the percentage of lime changed. The AEV of lime-treated soil decreased with increase in percentage of lime. This is due to flocculation, agglomeration into larger particles and chemical bonding, as a result of addition of lime, which create an open structure, resulting in larger void ratio The SWCC parameter "n", which is an approximate indicator of the pore size distribution and rate of desaturation, increases as the percentage of lime increases, indicating that lime-treated soils exhibit higher rate of desaturation than untreated natural soil.

Based on the results obtained in this study, it is recommended to carry out further research to investigate the potential of the use of S-SWCC to evaluate the improvement effect of stabilizers on expansive soils since S-SWCC is very effective in showing volume change effect. Such an approach may result in a better understanding of the optimum amount of lime content required for stabilization of expansive soils. Investigations on other factors influencing the SWCC function of treated soils such as initial water content, initial dry density, compaction effort, suction measuring methods, and hysteresis effects are also recommended.

CONFLICT OF INTEREST

The authors report there are no competing interests to declare.

ACKNOWLEDGEMENT

The first author acknowledges Bahirdar University for providing the funding for this study as part of the M.Sc. thesis study by the first author.

REFERENCES

[1] D. G. Fredlund, (2000). "The implementation of unsaturated soil mechanics into geotechnical engineering." *Canadian Geotechnical Journal,* 37, 963–986.

[2] M. Mavroulidou Z. Cabarkapa and M. J. Gunn (2013). "Efficient laboratory measurements of the soil water retention curve." *Geotechnical Test Journal* 36(1):1–9.

[3] C. W. W. Ng and Y. Pang (2000). "Experimental investigations of the soil–water characteristics of a volcanic soil." *Canadian Geotechnical Journal* 37:1252–1264.

[4] J. M. Tinjum and C. H. Benson (1997). "Soil–water characteristic curves for compacted clays." *J Geotechnical and Geoenvironmental Engineering* 123(11):1060–1069.

[5] S. K. Vanapalli D.E. Pufahl and D. G. Fredlund (1999). "The influence of soil structure and stress history on the soil–water characteristic of a compacted till." *Geotechnique, vol.* 49 no. 2, pp 143–159.

[6] X. Zhang, M. Mavroulidou, and M. J. Gunn, (2017). "A study of the water retention curve of lime-treated London Clay." *Acta Geotech.* 12, 23–45.

[7] D. Barman, and S. K. Dash, (2022). "Stabilization of expansive soils using chemical additives: A review." *Journal of Rock Mechanics and Geotechnical Engineering*, 14 (4), 1319–1342. https://doi.org/10.1016/j.jrmge.2022.02.011.

[8] F. G. Bell, (1988) "Lime stabilization of clay soils: part 1, basic principles." *Ground Eng* 21(1):10–15.

[9] M. Di Sante, E. Fratalocchi, F. Mazzieri, E Pasqualini, (2014) "Time of reactions in a lime treated clayey soil and influence of curing conditions on its microstructure and behavior." *Appl Clay Sci* 99:100–109.

[10] Z. Metelkova,´ J. Boha´cˇ. R. Prˇikryl and I. Sedla´rˇova,´ (2012) "Maturation of loess treated with variable lime admixture: pore space textural evolution and related phase changes." *Appl Clay Sci* 61:37–43.

[11] M. Cecconi, and G. Russo, (2013) "Microstructural features of lime stabilised pyroclastic soils." *Ge´otech Lett* 3(3):124–129.

[12] S. Khattab and L Al-Taie, (2006) "Soil–water characteristic curves (SWCC) for lime treated expansive soil from Mosul City." In: Unsaturated soils: 4th international conference on unsaturated soils (UNSAT 2006), ASCE, pp 1671–1682.

[13] ASTM D 421 – 85 (Reapproved 2002) "Standard Practice for Dry Preparation of Soil Samples for Particle-Size Analysis." Annual Book of ASTM Standards. - Philadelphia, PA: American Society for Testing Materials.

[14] H. Solomon, (2011). "Chemical Stabilization of Expansive Sub-grade Soil: Performance Evaluation on Selected Road Section in Northern Addis Ababa." MSc. Thesis, Addis Ababa University, Ethiopia.

[15] ASTM D6276 – 19 (2019). "Standard Test Method for Using pH to Estimate the Soil-Lime Proportion Requirement for Soil Stabilization" Annual Book of ASTM Standards. - Philadelphia, PA: American Society for Testing Materials.

[16] ASTM D6836 – 16 (2016). "Standard Test Methods for Determination of the Soil Water Characteristic Curve for Desorption Using Hanging Column, Pressure Extractor, Chilled Mirror Hygrometer, or Centrifuge." Annual Book of ASTM Standards. - Philadelphia, PA: American Society for Testing Materials.

[17] T. A. V. Gaspar, S.W. Jacobsz, M.V. Schulz-Poblete, and D. G. Toll, (2019). "Measurement of the soil water retention curve: practical considerations." Proceedings of the 17th African Regional Conference on Soil Mechanics and Geotechnical Engineering. 7, 8 & 9 October 2019 – Cape Town.

[18] D. G. Fredlund, (2019). "Determination of unsaturated soil property functions for engineering practice." Conference: 17th African Regional Conference on Soil Mechanics and Geotechnical Engineering At: Cape Town South Africa.

[19] Fredlund, D. G., and Xing, A. (1994). Equations for the Soil-Water Characteristic Curve. Canadian Geotechnical Journal. 31: 517-532.

[20] Fredlund, D. G., and Fredlund, M. D. (2020). Application of "Estimation Procedures" in Unsaturated Soil Mechanics. Geosciences, 10(9), 364.

[21] A. J. Puppala, K. Punthutaecha, and S. K. Vanapalli, (2006). Soil-Water Characteristic Curves of Stabilized Expansive Soils. Journal of Geotechnical and Geo environmental Engineering 132(6): 736–751.

Spectrum Occupancy Predictions Using Deep Learning Algorithms

Addisu Melkie Tafere[1] , Getachew Alemu Wondim[1*]

[1]*School of Electrical and Computer Engineering, Collage of Technology and Built Environment*
Addis Ababa University, Addis Ababa, Ethiopia.
[*]*Corresponding author's E-Mail address: getachew.alemu@aait.edu.et*

DOI: https://doi.org/10.63990/zede.v43i.12973

ABSTRACT

The fixed spectrum allocation (FSA) policy causes a waste of valuable and limited natural resources because a significant portion of the spectrum allocated to users is unused. With the exponential growth of wireless devices and the continuous development of new technologies demanding more bandwidth, there is a significant spectrum shortage under current policies. Dynamic spectrum access (DSA) implemented in a cognitive radio network (CRN) is an emerging solution to meet the growing demand for spectrum that promises to improve spectrum utilization, enabling secondary users (SUs) to utilize unused spectrum allocated to primary users (PUs). This study has addressed all the limitations of the previous studies by implementing a comprehensive approach that encompasses reliable spectrum sensing, potential candidate spectrum band identification, long-term adaptive prediction modeling, and quantification of improvements achieved in the prediction model. The Long-Short Term Memory (LSTM) Deep Learning (DL) model was proposed as a solution for this study to address the challenge of capturing temporal dynamics in sequential inputs. The LSTM model leverages a gating mechanism to regulate information flow within the network, allowing it to learn and model long-term temporal dependencies effectively. The dataset used for this study was obtained from a real-world spectrum measurement by employing the Cyclostationary Feature Detection (CFD) approaches in the GSM900 mobile network uplink band, spanning a frequency range of 902.5 to 915 MHz over five consecutive days. The dataset comprises a total of 225,000 data points. The five-day spectrum measurement data analysis yielded an average spectrum utilization of 20.47 %. The proposed model predicted the spectrum occupancy state for 5 hours ahead in the future with an accuracy of 99.45 %, improved the spectrum utilization from 20.47 % to 98.28 % and reduced the sensing energy to 29.39 % compared to real-time sensing.

Keywords: Cognitive radio spectrum, Deep learning, Dynamic spectrum access, Spectrum occupancy.

1. INTRODUCTION

The Radio Frequency (RF) spectrum is considered a limited and valuable natural resource used for various wireless communication systems, encompassing voice radio, digital terrestrial television (DTT), mobile telephony, and mobile broadband (MBB) [1]. The RF spectrum spans a wide range of electromagnetic waves demonstrating a direct relationship with their wavelength. Lower frequencies can propagate over longer distances and exhibit superior penetration through building walls. This characteristic makes them well-suited for applications such as broadcasting in expansive geographic areas. On the other hand, higher frequencies offer advantages in microelectronic devices like cell phones due to their shorter wavelengths that enable the use of proportionally smaller antennas, allowing these devices to transmit larger volumes of data [2].

As wireless technologies continue to advance, effective allocation and access remain essential for sustaining the growth

and reliability of wireless communication systems because they are limited resources that cannot be simultaneously used for different services due to interference. The demand for spectrum has increased dramatically due to the exponential growth of wireless devices and the continuous development of new bandwidth-hungry technologies. This has resulted in a spectrum scarcity, further increasing its commercial value. Spectrum stakeholders must, therefore, develop adaptable strategies to use the spectrum efficiently, meeting both current and future spectrum standards [1], [2], [3]. The current fixed spectrum allocation (FSA) policy is not addressing the growing demand for spectrum due to its rigid command-and-control approach, which assigns channels to a single user. This inefficient allocation can lead to the wastage of spectrum and a decrease in the quality of service. However, studies have shown that significant portions of the spectrum assigned to licensed users are unused, indicating the need for a more dynamic and responsive spectrum allocation policy [4], [5], [6].

Dynamic spectrum access (DSA) implemented in (CRNs) has emerged as a solution to improve spectrum utilization and reduce spectrum waste by allowing secondary users (SUs) to share unused portions with (PUs) [7]. CRNs comprise two types of users i.e., PUs and SUs, where PUs have a higher priority than SUs in accessing the channels. The SUs logically divides the channels allocated to the PUs into slots. Within each slot, the SUs has to sense the PU channel slot and accordingly access the slot when idle. The idle slots are called spectrum holes or white spaces [4]. This approach enhances spectrum utilization, accommodating the growing demand for wireless connectivity that enables more devices to be connected [8], [9]. Spectrum sharing requires knowledge of spectrum usage patterns, which can be obtained through spectrum sensing. However, real-time spectrum sensing is considered unreliable because of energy and time consumption. Spectrum occupancy state prediction is a technique that forecasts the future states of the spectrum proactively and estimates the effective bandwidth in the next slot allowing SUs to adjust their data rates in advance used to improve spectrum sensing. Therefore, SUs can conserve energy and time by avoiding the busy portions of the spectrum and focusing on idle portions during sensing [4], [7]. Spectrum occupancy state prediction which infers the future states of the spectrum channel, is a key enabler for shared spectrum access in the DSA model. Proactive spectrum prediction allows SUs to identify and access idle spectrum channels before they become busy [7], [10], [11]. SUs in CRNs search for idle spectrum channels to use temporarily. They are equipped with the cognitive ability to effectively implement the CR, which performs the following cycle of functions: Sensing: to observe and sample spectral channels, Decision: to allocate suitable spectral holes, Sharing: to contend access with other SUs, and Mobility: to evacuate the spectral hole when a PU is present [11]. CRs have distinct characteristics that distinguish them from traditional radio systems and Software-Defined Radios (SDRs). These distinctive characteristics are a cognitive capability that enables the identification of the occupancy state and usage patterns of the spectrum channels and reconfigurability which allows them to adjust their operating parameters dynamically [12], [13], [14]. CRNs are intelligent networks capable of autonomously learning and dynamically adapting to optimize spectrum, which is inherent in the adaptability of DL models that excel in learning complex patterns from data to make informed decisions. The integration of DL models into CRNs with the capacity to analyze vast amounts of data and enhance the awareness of CRNs about their operating environment holds significant potential in spectrum-sharing

models. The DL models can adapt and learn from the changing spectrum state conditions, allowing CRNs to dynamically optimize their communication parameters, spectrum channels, and transmission scenarios. Therefore, the synergy presents a relationship where DL's learning capabilities empower CRNs with enhanced situational awareness for intelligent decision-making [15], [16].

A distributed spectrum management framework for mobile edge computing (MEC)-based Cognitive Radio Internet of Things (CR-IoT) networks was proposed to integrates edge computing that enhance real-time decision-making and reduce latency in spectrum allocation [17]. Using a game-theoretic and optimization-based approach, it enables efficient and autonomous spectrum sharing among IoT devices while minimizing interference. Simulation results demonstrate improved spectrum utilization, energy efficiency, and network throughput compared to traditional centralized spectrum management methods.

A joint resource allocation and user association framework was proposed for multi-cell Integrated Sensing and Communication **(**ISAC) dense networks [18]. Interference models for sensing and communication were established, and a utility-maximization problem was formulated under SINR and power constraints. A greedy genetic sub-band allocation, Hungarian-based user association, and SCA-based power control were used to solve the non-convex problem. Simulations showed notable improvements in network utility and detection probability while balancing sensing–communication trade-offs.

Recent studies on machine learning-based spectrum occupancy prediction focus on improving the efficiency of spectrum utilization in CRNs [19]. Techniques such as SVM, Artificial Neural Networks (ANN), CNN, Recurrent Neural network (RNN), and ensemble models have been widely applied to predict temporal and spatial spectrum usage patterns. Researchers emphasize that deep learning models outperform traditional statistical methods in capturing nonlinear and dynamic spectrum behaviors. However, challenges remain in data scarcity, real-time prediction, and generalization across frequency bands**.**

This study has achieved promising results in solving the limitations of previous studies. In general, the main contributions of this study are:

1. Defining the PU channel characterization in a new mode in a time-domain approach called CFD to characterize primary user states.
2. Identifying the potential candidate spectrum band for a CR deployment through proper spectrum utilization and techno-economic analysis.
3. Developing an improved long-term spectrum occupancy state prediction that can predict the spectrum occupancy state of how long it will be busy and idle, which allows the SUs to improve spectrum access, reduce channel-switching costs, and increase the CRN throughput.
4. Quantifying the improvements achieved in the spectrum occupancy state prediction model.

2. STATEMENT OF THE PROBLEM

In spectrum sensing, parametric approaches rely on prior information about (PU) activity, whereas in many real-world cases, such information is unavailable. Consequently, nonparametric sensing methods, particularly Energy Detection (ED), are widely used due to their low computational complexity and ease of implementation [20]. However, the wireless environment introduces issues such as fading and hidden node problems, causing an exponential decay of field strength during transmission. This makes threshold selection difficult at low signal-

to-noise ratios (SNRs), rendering ED inefficient and interference-prone in Cognitive Radio (CR) systems [8], [9], [13]. In contrast, the parametric sensing approach known as (CFD) outperforms ED because it exploits the spectral correlation of cyclostationary signals—a property absent in noise. This allows CFD to operate effectively in low SNR regions, remain robust to noise uncertainty, and deliver superior performance when prior information about the PU signal is available [13], [14].

Traditional statistical models such as Autoregressive Integrated Moving Average (ARIMA), Hidden Markov Model (HMM) and Markov Chains and machine learning algorithms assume spectrum occupancy states as stationary processes, implying that they remain constant over time and are suitable for short-term predictions [6], [21]. ANNs are less effective for modeling temporal data due to the absence of memory elements. RNNs have been employed for such tasks; however, they face challenges, such as the vanishing gradient problem, hindering their ability to capture long-term dependencies [7], [20]. LSTM neural networks have been introduced to address the vanishing problem. LSTMs overcome the vanishing gradient problem by incorporating memory cells, allowing them to retain information over extended periods. This feature is particularly advantageous for modeling temporal data, such as spectrum occupancy [15], [20], [21]. Long-term spectrum occupancy state prediction plays a crucial role in anticipating the channel idle period duration, which reduces channel switching costs and enhances channel selection in CRNs.

3. SYSTEM MODEL

The heterogeneous spectrum occupancy state model was used to implement a CRN for spectrum sharing in a DSA model [15]. The spectrum band has been divided into k contiguous channels. The channel states represent the spectrum channel state at time t and are denoted by a vector matrix. Each element in the matrix represents the corresponding channel is occupied or idle which is ready to be used by the SUs [22]. A heterogeneous spectrum occupancy state model has multiple PUs and SUs, which are centrally controlled by a database that identifies the spectrum occupancy states based on prediction [4]. This study has proposed a long-term spectrum occupancy state prediction model in a stationary location by exploiting the spectral and temporal correlation of the data. The occupancy of a channel is characterized by the presence of a primary user signal, while the presence of a spectrum hole characterizes the vacancy of a channel. These cases are formally stated as hypotheses (H0) and (H1).

$$H_0: y[t] = w[t] \quad \text{when there is no PU}$$
$$H_1: y[t] = h[t]x[t] + w[t] \quad \text{when PU's signal is present} \qquad (1)$$

where x[t] denotes the PU signal, w[t] is white noise and y[t] is the received signal at t^{th} time instant. H_0, the null hypothesis indicates the noise samples while H_1, the alternate hypothesis indicates the presence of PU signal along with noise t^{th} instant [6], [8], [12].

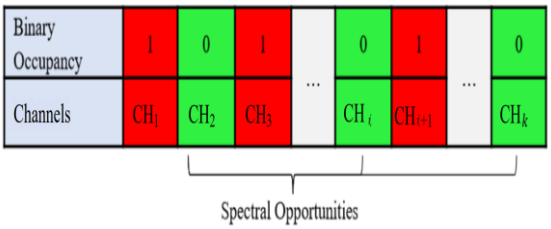


Figure 1 Spectrum channels occupancy state modeling

For the sequentially obtained time-series spectrum occupancy measurement data X_1, X_2, X_3, X_4,..., the long-term spectrum channel state prediction using the deep learning model can be done within the sequence-to-sequence neural network architecture based on the LSTM deep learning model is defined as (X_{t-n},..., X_{t-2},

$X_{t-1)\ to}$ (X_t, X_{t+1},…., $X_{t+m)}$ where *n* and *m* represents the historical observations and the future instants in time, respectively [23].

3.1 Time-Series Model Analysis

Time series data is composed of sequence data points measured over regular intervals. Time series models are designed to comprehend the patterns and trends inherent in the data that can feat the natural temporal ordering. This shows that observations closer in time are more similar than further apart observations because values in a time series data at a given time were derived from past values. Time series analysis is a methodology used to study time series data that identifies relationships and makes predictions of future values [24]. Time series data analysis begins by identifying whether the data is stationary or non-stationary. A stationary time series has consistent statistical properties, including mean, variance, and autocorrelation. This stability implies that the underlying data-generating process is predictable and stable. On the contrary, non-stationary time series data displays varying statistical properties over time, often influenced by trends, seasonality, or other non-random fluctuations. Time series prediction is a technique that estimates future values based on historical data. In spectrum occupancy state prediction, the objective is to forecast the spectrum occupancy state at the next time point [24], [25].

Time series data should be selected carefully, considering the different variations that can occur at different timescales. For example, a day can have four seasons (morning, afternoon, evening, and night), while a week can have only two (weekday and weekend). The spectrum occupancy data can vary significantly depending on the time of day, day of the week, and peak or trough times. Therefore, representative data must be chosen for the specific period and timescale of interest [26].

3.2 Privacy Concerns of Urban Spectrum Sensing

Urban spectrum sensing involves the collection and analysis of spectrum usage data from multiple SUs distributed across densely populated areas. While such collaborative or crowdsourced sensing improves the accuracy of spectrum occupancy prediction and the efficiency of spectrum utilization, it also introduces significant privacy concerns [27]. These arise primarily from the collection of user-related information that may inadvertently reveal sensitive details about a user's location, behavior, and communication patterns [27], [28].

Sources of Privacy Risks

i) **Location Disclosure:** Spectrum sensing data often includes information about the spatial position of participating SUs. Since sensing reports are typically associated with the geographical coordinates of the sensors, malicious entities or untrusted fusion centers can infer the exact location or movement patterns of users [29].

ii) **Activity Inference:** The frequency and timing of sensing reports can reveal user activities or communication habits. For example, consistent sensing from a specific location or time interval can indicate when a user is active, which networks they are connected to, or even which applications are being used [27].

iii) **Data Correlation and Identification:** Aggregated sensing data from multiple users may be correlated to re-identify individuals even when identifiers are removed. Advanced data mining or machine learning techniques can exploit these correlations to infer private user attributes [30].

iv) **Malicious Data Collection:** In cooperative sensing scenarios, untrusted or compromised nodes can collect data not only for spectrum

purposes but also for unauthorized surveillance, traffic analysis, or profiling [28], [31].

Privacy Preservation Mechanisms

To mitigate these risks, several privacy-preserving spectrums sensing schemes have been proposed. These mechanisms aim to balance data utility with user confidentiality, ensuring that accurate spectrum decisions can still be made without compromising user privacy.

i) **Data Anonymization:** This technique removes or obscures identifiable information before transmitting sensing data to the fusion center. However, anonymization alone is often insufficient, as de-anonymization attacks can exploit contextual data [32].

ii) **Differential Privacy:** Differential privacy adds controlled random noise to sensing data, preventing adversaries from accurately inferring individual user contributions. It provides quantifiable privacy guarantees while maintaining statistical data utility [33].

iii) **Cryptographic Techniques: Homomorphic encryption** allows users to encrypt their sensing data so that the fusion center can perform computations without accessing the raw data, and secure **multiparty computation (SMC)** enables distributed users to collaboratively compute spectrum availability without revealing their individual inputs [28], [32].

iv) **Decentralized and Edge-Based Sensing:** Instead of central data aggregation, edge computing approaches process data locally on user devices or local base stations, transmitting only aggregated or decision-level information. This significantly reduces the exposure of raw sensing data [30], [32].

3.3 Proposed Deep Learning LSTM Architecture

The proposed model can accurately predict the spectrum occupancy state for the next time slot and several time slots ahead, implemented to facilitate a DSA. This model enables users to access spectrum channels that are not used by PUs, enhancing overall spectrum utilization efficiency. (Figure 2).

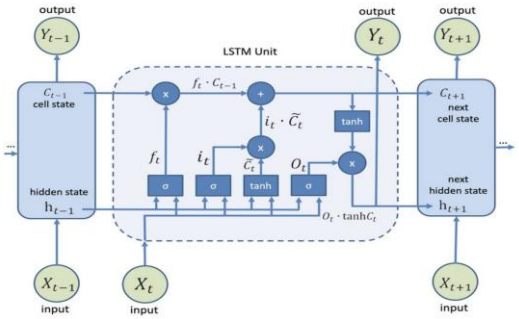


Figure 2 Basic Architecture of the LSTM model

The LSTM model used for long-term spectrum occupancy state prediction targets specific frequencies or channels, using known binary values for state estimation. Input and output data are constructed via a sliding window across both time and frequency axes, forming a 2D matrix where each element represents a time–frequency point with its binary state. This matrix constitutes the training dataset. During validation, the model demonstrates real-time prediction capability, using past spectrum measurements over time and frequency lags to predict the next state. The binary grid input yields corresponding binary outputs for the subsequent instant [34]. Unlike conventional RNNs, which struggle with long-range dependencies due to the vanishing gradient problem, LSTMs effectively model such temporal dynamics. LSTM networks address the problem of long-term dependency by introducing a gating mechanism that regulates the flow of information within the network. The mechanism consists of three gates the input, forget, and output gates and a cell

memory used to retain information over long periods [7], [15].

1) **Input gate:** Controls how much new information is added to the LSTM's memory state.

2) **Forget gate:** Determines how much previous information is discarded.

3) **Output gate:** Regulates how much of the cell memory is passed to the next hidden state. The gates are mathematically formulated as:

$$i_t = \sigma\left(W_{ix}X_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i\right)$$

$$f_t = \sigma\left(W_{fx}X_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f\right)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \varphi\left(W_{cx}X_t + W_{ch}h_{t-1} + b_c\right)$$

$$o_t = \sigma\left(W_{ox}X_t + W_{oh}h_{t-1} + W_{oc}c_t + b_0\right)$$

$$h_t = o_t \circ \varphi\left(c_t\right) \quad (2)$$

Where i, f, o, and c denote the input gate, forget gate, output gate, and cell state, respectively. Each gate has the same dimension as the hidden vector h (N ×1). Here σ is a sigmoid function, and φ is a nonlinear function mapping to the range [-1, 1]. W_{ic}, W_{fc}, and W_{oc} are the peephole connection matrices linking the cell state to their respective gates, while W_{ix}, W_{fx}, W_{ox}, and W_{cx} are input weight matrices connecting the input vector X_t (M × 1) to the gates and cell state. Because of the gates and the input vector X_t have the dimensions of N ×1 and M × 1 respectively. The matrices W_{ih}, W_{ic}, W_{fh}, W_{ch}, W_{oh}, W_{oc} are all of dimensions N × N, and W_{ix}, W_{fx}, W_{cx} and W_{ox} are of dimensions N × M.

4. RESULTS AND DISCUSSION

4.1 Data Description

The data used for this study was collected from a real-world spectrum measurement in Addis Ababa, Ethiopia using the TCI spectrum monitoring system in the GSM900 MHz mobile network uplink band spanning from 902.5 to 915 MHz for five consecutive days from January 28th to February 1st, 2021. The area of Bole was selected for the measurement because it is a commercial area that expects to have a high spectrum demand [35]. The data set comprised a total of 450,000 data points, captured with a resolution bandwidth of 100 kHz with 4 minutes resolution time. The GSM900 MHz uplink band was selected for this study to deploy a CR due to its underutilization from the sparse use of its users communicating on the network making it a promising potential candidate for a CR deployment [36], [37].The dataset contains features including Channel, Frequency, Maximum occupancy (%), Average occupancy (%), Maximum field strength, and Average field strength. However, all these features were not used for modeling spectrum occupancy due to their inability to capture unique information and their potential negative impact on the model's generalization. To address this issue, a feature reduction technique was employed to reduce the number of input variables, thus preventing excessive model complexity while preserving its ability to generalize effectively. In this study, a filter-based feature selection approach was utilized, which relies on statistical measures such as information gain, to identify features that contribute the most information about the target variable which exclusively considers the association between each feature and the class label [38]. Following the feature reduction process, the selected features for predicting spectrum occupancy state comprised frequency, average occupancy, and average field strength.

The spectrum measurement campaign that was conducted in the GSM900 MHz mobile network uplink band at four different regional cities in October 2021 has an average utilization of 21.45 % in Adama, 16.21 % in Bahir Dar, 18.87 % in

Hawassa, and 33.52 % in Jigjiga. Additionally, within Ethio-telecom's mobile network uplink bands in Addis Ababa, utilization was found to be 14.72 % in GSM 900MHz, 31.67 % in UMTS 900MHz, 25.32 % in LTE 1800MHz, and 6.75 % in LTE 2600MHz. The analysis indicates the underutilization of the spectrum. Even though all of the spectrum bands are underutilized, subscribers would migrate from 2G to 3G, 4G, and 5G to get the advanced technology features and services from these new-generation mobile networks because the 900MHz GSM offers only voice and short message services (SMS) [39]. For this reason, there is no need of considering 3G, 4G/LTE and 5G networks in our work.

The spectrum utilization analysis conducted in five days revealed distinct values for weekdays (Thursday, Friday, and Monday) and weekends (Saturday and Sunday). Specifically, the average spectrum utilization on weekdays was 19.13 %, 19.03 %, and 21.14 %, respectively. In contrast, the average spectrum utilization on weekends exhibits a variation, with values of 17.3 % on Saturday and a higher utilization rate of 25.76 % on Sunday.

The CFD-based spectrum sensing method was used in this study for defining PU channel characterization and modeling spectrum occupancy prediction due to its enhanced performance in challenging SNR conditions. In a five-day spectrum measurement campaign conducted using the CFD spectrum sensing method, the average spectrum utilization for the GSM 900 MHz uplink band was 20.47 %.

4.2 Results

This study utilizes data obtained via the CFD-based spectrum sensing method to develop the spectrum occupancy state prediction model. The dataset encompasses 225,000 data points, representing half of the five-day measurement dataset. Subsequently, this dataset is divided into training and validation sets, maintaining an 80 % to 20 % ratio, we should notice that this is done to minimize the risk of overstated generalizability. The spectrum occupancy state prediction model was implemented using Python programming with the Keras library. The model's configuration was evaluated based on metrics such as the loss function and accuracy (Figure 3). Throughout the training process, various hyperparameter combinations were explored to identify the most effective model that has the lowest loss and the highest accuracy. Finally, the model architecture comprises three LSTM layers with 128 units, followed by two dropout layers with a dropout rate set at 0.1. Additionally, two dense layers with 128 and 64 units, were incorporated, along with a final output layer. The model was configured with an activation function of rectified linear unit (ReLU) for all hidden layers and sigmoid for the output layer, adaptive moment estimation (ADAM) as an optimizer, binary cross-entropy as a loss function, 0.001 learning rate, 128 batch size, and 400 Epochs. Three sets of experiments designed to predict short-term and long-term (ranging from three to five hours) predictions conducted on the proposed LSTM model have exhibited consistent results across all performance evaluation metrics with an accuracy of 99.45 %.

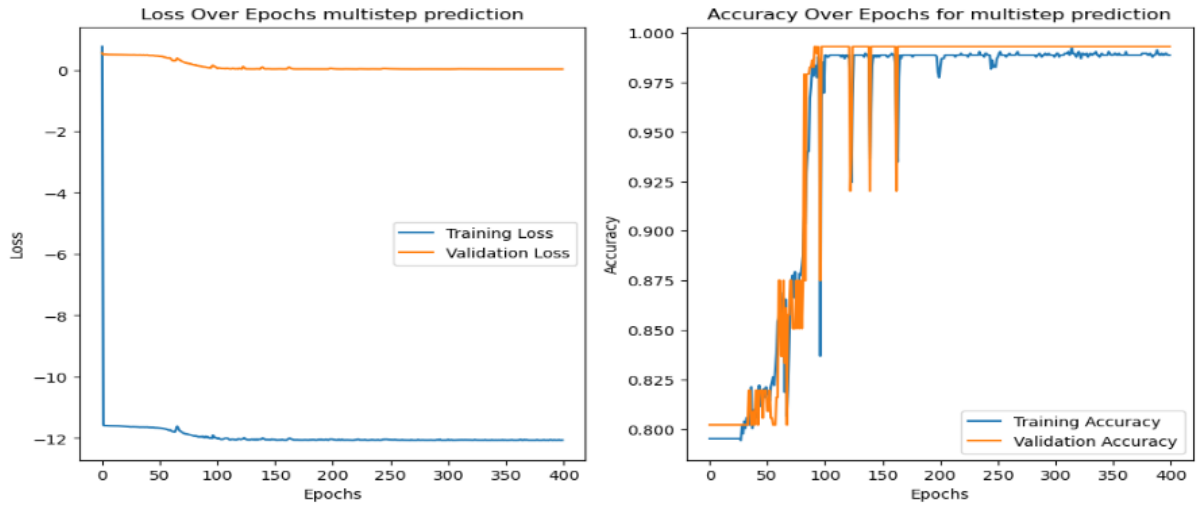


Figure 3 Training and validation, loss and accuracy for the LSTM model for 5-hours ahead prediction

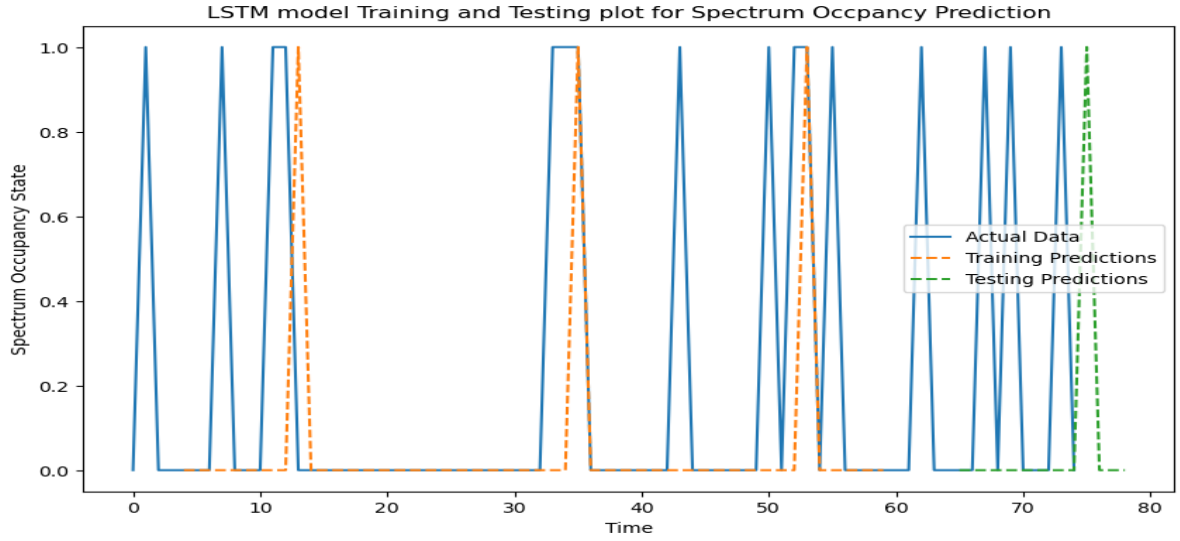


Figure 4 Spectrum occupancy state prediction in LSTM for the 904.1MHZ channel

Figure 4 compares the spectrum occupancy state actual data, training, and testing prediction performed for the 904.1MHZ channel. The LSTM network is trained to adapt new spectrum occupancy states, as shown in Figure 4, using the five-day spectrum measurement data for a one-channel 904.1MHz. In the graph, the brown dotted line representing the training data and the green dotted line representing the testing data closely match the actual observations depicted by the blue solid line. The test performance indicates an accuracy of 96 %.

A comparative analysis conducted on LSTM, Bi-LSTM, and ConvLSTM models for both short-term and long-term prediction revealed that LSTM and Bi-LSTM achieved equal results, with an accuracy of 99.45 %. In contrast, the ConvLSTM model outperforms them, achieving an accuracy of 99.72 %. Across all three models, each term of prediction has equal performance results. However, the short-term prediction achieved its targeted accuracy earlier than the long-term 3-hour and 5-hour predictions. The 3-hour long-term prediction achieved its targeted accuracy earlier than the 5-hour long-term prediction.

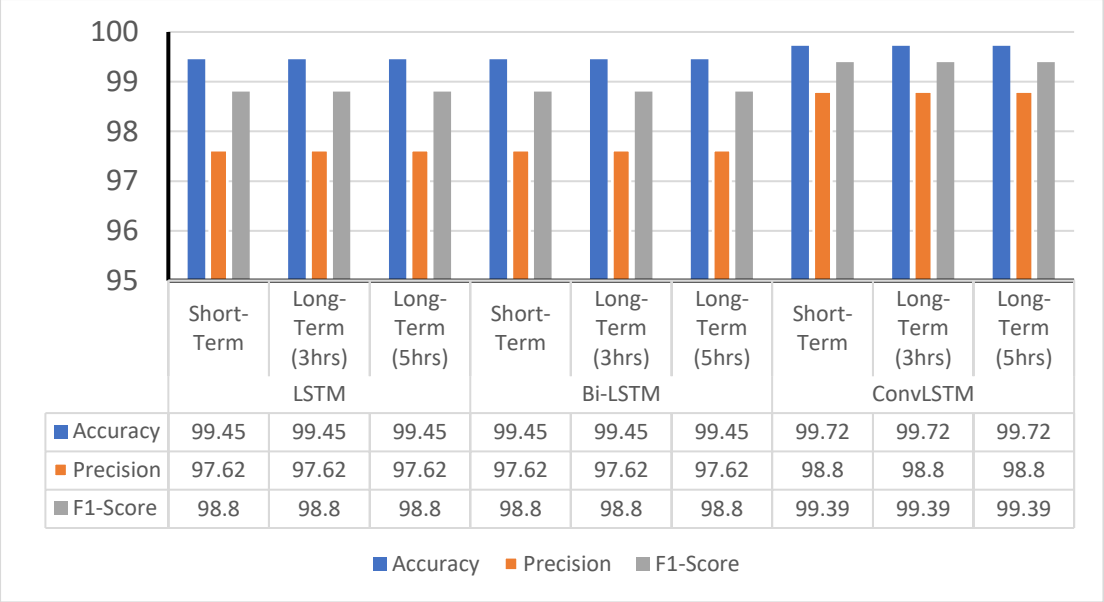


Figure 5 Comparison of models in Accuracy, Precision, and F1-Score

The Bi-LSTM model differs from the others by achieving its targeted accuracy earlier across all prediction scenarios. This may be attributed to its ability to process data in both forward and backward directions, potentially improving its ability to learn temporal relationships. Consequently, short-term predictions tend to achieve better accuracy earlier than long-term predictions. This tendency may arise from factors such as predicting the immediate future requires fewer data points and simpler relationships, and the available data for training the model can be enough for short-term prediction than for long-term prediction.

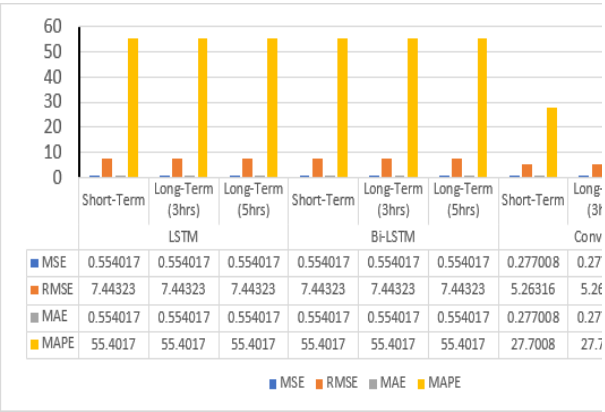


Figure 6 Comparison of models in MSE, RMSE, MAE, and MAPE

4.3 Discussions

Long-term spectrum occupancy state prediction using the LSTM model to implement a CR has improved spectrum utilization and reduced sensing energy.

Improvement in Spectrum Utilization

Spectrum occupancy state prediction can improve spectrum utilization by allowing SUs to select and use idle PU channels with appropriate time slots. This allows the SUs to select spectrum channels efficiently to reduce channel-switching costs and increase network throughput. In a CRN the PUs has two states, but the SU can sense only one channel at a time. The CRN has two types of SUs which are the CRsense and the CRpredict. The CRsense randomly selects a channel at every slot and senses the states of the channels, while the CRpredict device senses the states of the channels after prediction among those channels with an idle state. According to [4] spectrum utilization (SU) in the CRN can be defined as the ratio of the number of idle slots discovered by the SUs to the total number of idle slots available in the CRN.

$$SU = \frac{\text{Number of idle slots sensed}}{\text{Total number of idle slots in the band}} \quad (3)$$

The improvement in spectrum utilization due to spectrum prediction can be expressed as

$$SU_{imp(\%)} = \frac{SU_{sense} - SU_{predict}}{SU_{sense}} \quad (4)$$

Where SU_{sense} and $SU_{predict}$ represent the spectrum utilization for the CR_{sense} and the $CR_{predict}$ devices, respectively. Substituting (3) in (4), $SU_{imp(\%)}$ can be given by

$$SU_{imp(\%)} = \frac{I_{sense} - I_{predict}}{I_{sense}} \quad (5)$$

Where I_{sense} and $I_{predict}$ represent the number of idle channels sensed by the CR_{sense} and the number of idle channels predicted by the $CR_{predict}$ devices respectively. This analysis can be translated into a machine learning model and becomes equal with the specificity that measures the true negative rate, which is the fraction of negative values that were correctly predicted that can be calculated as expressed in (6).

$$SU_{imp(\%)} = Specifcity = \frac{TN}{TN+FP} \quad (6)$$

The $SU_{imp(\%)}$ in a CRN also improves the network throughput which shows the data rate achieved in the network due to the availability of more channels and can be calculated as expressed in (7).

Throughput = $SU_{imp(\%)}$* the number of channels in the spectrum band (7)

Reduction in Sensing Energy

The spectrum occupancy state prediction reduces the sensing energy required by the SUs. This is because the SUs in a CRN can sense only the idle channels. In a CRN the CR_{sense} device senses all the channels whereas the $CR_{predict}$ device only senses the channel that is predicted idle.

In other words, when the channel state is predicted to be busy, the sensing operation is not performed to reduce energy. Considering that one unit of sensing energy is required to sense one slot, the total sensing energy required for a CR_{sense} device in a finite duration of time can be calculated as expressed in (8).

$$SE_{sense} = \begin{pmatrix}\text{Toatal number}\\ \text{of slots}\\ \text{in the duration}\end{pmatrix} \times \begin{pmatrix}\text{unit sensing}\\ \text{energy}\end{pmatrix} \quad (8)$$

while the total sensing energy required by the $CR_{predict}$ device can be given by

$$\begin{aligned}&\text{SE}_{\text{predict}}\\ &= \left(\text{SE}_{\text{sense}} - \left(\text{B}_{\text{preidct}}\right)\right)\\ &* \ (\text{Unit sensing energy})\end{aligned} \quad (9)$$

Where $B_{preidct}$ is the total number of busy slots predicted by the $CR_{predict}$ device.

Therefore, using (8) and (9), the percentage reduction in the sensing energy can be given by

$$\begin{aligned}SE_{red(\%)} &= \frac{SE_{sense} - SE_{predict}}{SE_{sense}}\\ &= \frac{B_{predict}}{Total\ no.of\ idle\ slots}\end{aligned} \quad (10)$$

This can be translated to a machine learning model that measures a value by dividing the true positive value by the true negative plus the false positive values, even it doesn't have an equivalent machine learning metric it can be calculated and expressed as shown in (11).

It should be noted that Equation (11) is derived from analytical formulations reported in the literature. This study extends these formulations by quantifying sensing energy reduction using machine learning metrics. The resulting expression links the analytical energy model to classification outcomes—true positives (TP), true negatives (TN), and false positives (FP). While conceptually similar to standard ML metrics, it is **not** equivalent to precision or recall. Accordingly, we define this parameter as representing the proportion of sensing energy saved through accurate spectrum state prediction**.**

$$SE_{red(\%)} = \frac{TP}{TN+FP} \quad (11)$$

The $_{SUimp\ (\%)}$, the network throughput, and the $_{SEred\ (\%)}$ across all models for all terms of predictions calculated based on (6), (7), and (11) are presented in Table 1.

Table 1 Quantified improvements achieved in the spectrum occupancy state prediction

Model	**Length of Prediction**	$SU_{imp(\%)}$	**Throughput**	$SE_{red(\%)}$
LSTM	Short-Term	99.28	124.1	29.39
	Long-Term (3 hrs.)	99.28	124.1*45	29.39
	Long-Term (5 hrs.)	99.28	124.1*75	29.39
Bi- LSTM	Short-Term	99.28	124.1	29.39
	Long-Term (3 hrs.)	99.28	124.1*45	29.39
	Long-Term (5 hrs.)	99.28	124.1*75	29.39
Conv-LSTM	Short-Term	99.64	124.55	29.39
	Long-Term (3 hrs.)	99.64	124.55*45	29.39
	Long-Term (5 hrs.)	99.64	124.55*75	29.39

5. CONCLUSION AND FUTURE WORK

This study addresses the challenges posed by the rigidity of FSA policies. It paves the way for more effective and efficient spectrum utilization that optimizes scarce spectrum resources by predicting the spectrum occupancy state using a long-term adaptive LSTM model and a historical dataset obtained through a reliable spectrum sensing method. The proposed model has successfully predicted the spectrum occupancy state for the subsequent five hours with an accuracy of 99.45 % improved the spectrum utilization from 20.47 % to 98.28 % and reduced the sensing energy to 29.39 % compared to real-time sensing. Future studies can focus on enhancing the predictability of further occupancy lengths up to days for integrating CRN with the Internet of Things, which creates a synergistic system known as the (CRIoT). This integrated IoT and CR approach amplifies smart cities' capability, providing a comprehensive and interconnected infrastructure for effective and efficient urban management.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest in this work.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Addis Ababa University, Addis Ababa Institute of Technology, School of Electrical and Computer Engineering. The accomplishment of the paper would not have been possible without its support.

REFERENCES

[1] J. Kalliovaara, "Field Measurements in Determining Incumbent Spectrum Utilization and Protection Criteria in Wireless Co-existence Studies," *Turku University of Applied Sciences,* 2017. [Online]. Available: DOI:10.13140/RG.2.2.33351.80805

[2] F. Weidling, D. Datla, V. Petty, P. Krishnan, and G. J. Minden, "A Framework for R.F. Spectrum Measurements and Analysis," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, 2005. [Online]. Available: DOI: 10.1109/DYSPAN.2005.1542672.

[3] T. Jeacock, "A Standard Approach for Assessing the Spectrum Management Needs of Developing Countries," Report on Telecommunication Development Sectror: ITU, 2016. [Online]. Available: www.itu.int.

[4] V. K. Tumuluru, P. Wang, and D. Niyato, "A Neural Network Based Spectrum Prediction Scheme for Cognitive Radio," in *IEEE International Conference on Communications*, 2010. doi: 10.1109/ICC.2010.5502348.

[5] L. J. H. Viveros, D. A. L. Sarmiento, and N. E. V. Parra, "Modeling and Prediction Primary Nodes in Wireless Networks of Cognitive Radio Using Recurrent Neural Networks," *Contemporary Engineering Sciences*, 2018, doi: 10.12988/ces.2018.84164.

[6] N. G. Bara'u, W. Feng, and M. Almustapha, "Spectrum Hole Prediction Based on Historical Data: A Neural Network Approach," *Computer Science, Engineering*, 2014, [Online]. Available: Corpus ID: 6991758.

[7] B. S. Shawel, D. H. Woledegebre, and S. Pollin, "Deep-learning based Cooperative Spectrum Prediction for Cognitive Networks," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 2018.

[8] M. Sardana and A. Vohra, "Analysis of Different Spectrum Sensing Techniques," in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, 2017. [Online]. Available: DOI: 10.1109/COMPTELIX.2017.8004006.

[9] P. S. Yawada and A. J. Wei, "Comparative Study of Spectrum Sensing Techniques Base on Techniques Non-cooperative in Cognitive Radio Networks," in *2016 5th International Conference on Computer Science and Network Technology (ICCSNT)*, 2016. [Online]. Available: DOI: 10.1109/ICCSNT.2016.8070212.

[10] A. Saad, B. Staehle, and R. Knorr, "Spectrum Prediction using Hidden Markov Models for Industrial Cognitive Radio," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016. [Online]. Available: DOI: 10.1109/WiMOB.2016.7763231.

[11] T. Wysocki and B. J. Wysocki, "Spectrum Occupancy Prediction Using a Hidden Markov Model," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016. Accessed: Mar. 12, 2024. [Online]. Available: DOI: 10.1109/ICSPCS.2015.7391772.

[12] N. Muchandi and R. Khanai, "Cognitive radio spectrum sensing: A survey," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016. [Online]. Available: DOI: 10.1109/ICEEOT.2016.7755301.

[13] M. Z. Alom, T. K. Godder, and M. N. Morshed, "A Survey of Spectrum Sensing Techniques in Cognitive Radio Network," in *2015 International Conference on Advances in Electrical Engineering (ICAEE)*, 2015. [Online]. Available: DOI: 10.1109/ICAEE.2015.7506821

[14] V. Amrutha and K. V. Karthikeyan, "Spectrum Sensing Methodologies in Cognitive Radio Networks: A Survey," in *2017 International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT)*, 2017. [Online]. Available: DOI: 10.1109/ICIEEIMT.2017.8116855.

[15] A. Shenfield, Z. Khan, and H. Ahmadi, "Deep Learning Meets Cognitive Radio: Predicting Future Steps," in 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020. [Online]. Available: http://shura.shu.ac.uk/25962.

[16] B. Soni, D. K. Patel, and M. L.Benitez, "Long Short-Term Memory based Spectrum Sensing Scheme for Cognitive Radio using Primary Activity Statistics," *IEEE Access,* 2020. doi: 10.1109/ACCESS.2017.

[17] M. A. Aygül, H. A. Çırpan, and H. Arslan, "Machine learning-based spectrum occupancy prediction: a comprehensive survey," *Frontiers in Communications and Networks*, vol. 6, no. January, pp. 1–21, 2025, doi: 10.3389/frcmn.2025.1482698.

[18] J. Zhang, Z. Fei, X. Wang, P. Liu, J. Huang, and Z. Zheng, "Joint resource allocation and user association for multi-cell integrated sensing and communication systems," *EURASIP J Wirel Commun Netw*, vol. 2023, no. 1, Dec. 2023, doi: 10.1186/s13638-023-02264-1.

[19] J. Wang and J. Wang, "A New Hybrid Forecasting Model Based on SW-LSTM and Wavelet Packet Decomposition: A Case Study of Oil Futures Prices," *Comput Intell Neurosci*, vol. 2021, 2021, doi: .10.1155/2021/7653091.

[20] N. Balwani, D. K. Patel, B. Soni, and M. L. Benítez, "Long Short-Term Memory based Spectrum Sensing Scheme for Cognitive Radio," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019. [Online]. Available: DOI: 10.1109/PIMRC.2019.8904422.

[21] M. A. Aygül, M. Nazzal, M. İ. Sağlam, D. B. da Costa, H. F. Ateş, and H. Arslan, "Efficient spectrum Occupancy Prediction Exploiting Multidimensional Correlations through Composite 2D-LSTM Models," *Sensors (Switzerland)*, 2020, doi: 10.3390/s21010135.

[22] L. Yu, Q. Wang, Y. Guo, and P. Li, "Spectrum Availability Prediction in Cognitive Aerospace Communications: A Deep Learning Perspective," in *2017 Cognitive Communications for Aerospace Applications Workshop (CCAA)*, IEEE, 2017. [Online]. Available: DOI: 10.1109/CCAAW.2017.8001877.

[23] B. S. Shawel, D. H. Woldegebreal, and S. Pollin, "Convolutional LSTM-based Long-Term Spectrum Prediction for Dynamic Spectrum

Access," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019. [Online]. Available: DOI: 10.23919/EUSIPCO.2019.8902956.

[24] Z. Wang and S. Salous, "Spectrum Occupancy Statistics and Time Series Models for Cognitive Radio," *J Signal Process Syst*, Feb. 2011, doi: 10.1007/s11265-009-0352-5.

[25] A. Agarwal, A. S. Sengar, and R. Gangopadhyay, "Spectrum Occupancy Prediction for Realistic Traffic Scenarios: Time Series versus Learning-Based Models," *Journal of Communications and Information Networks*, vol. 3, no. 2, Jun. 2018, doi: 10.1007/s41650-018-0013-6.

[26] L. Yin, S. Yin, W. Hong, and S. Li, "Spectrum Behavior Learning in Cognitive Radio based on Artificial Neural Network," in *2011 - MILCOM 2011 Military Communications Conference*, 2011. [Online]. Available: DOI: 10.1109/MILCOM.2011.6127671

[27] A. Javed *et al.*, "Review of spectral indices for urban remote sensing," *Photogramm Eng Remote Sensing*, vol. 87, no. 7, pp. 513–524, 2021, doi: 10.14358/PERS.87.7.513.

[28] M. Wasilewska, H. Bogucka, and H. V. Poor, "Secure Federated Learning for Cognitive Radio Sensing," *IEEE Communications Magazine*, vol. 61, no. 3, pp. 68–73, Mar. 2023, doi: 10.1109/MCOM.001.2200465.

[29] S. E. Abdelbaset, H. M. Kasem, A. A. Khalaf, A. H. Hussein, and A. A. Kabeel, "Deep Learning-Based Spectrum Sensing for Cognitive Radio Applications," *Sensors*, vol. 24, no. 24, Dec. 2024, doi: 10.3390/s24247907.

[30] J. Zhang, Z. Fei, X. Wang, P. Liu, J. Huang, and Z. Zheng, "Joint resource allocation and user association for multi-cell integrated sensing and communication systems," *EURASIP J Wirel Commun Netw*, vol. 2023, no. 1, Dec. 2023, doi: 10.1186/s13638-023-02264-1.

[31] M. U. Muzaffar and R. Sharqi, "A review of spectrum sensing in modern cognitive radio networks," Feb. 01, 2024, *Springer*. doi: 10.1007/s11235-023-01079-1.

[32] I. E. Ezhilarasi, J. C. Clement, and J. M. Arul, "A survey on cognitive radio network attack mitigation using machine learning and blockchain," Dec. 01, 2023, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1186/s13638-023-02290-z.

[33] M. Schukat, E. Barrett, M. Morales Cespedes, and A. G. Armada, "Enhanced spectrum sensing for AI-enabled cognitive radio IoT with noise uncertainty," 2025. [Online]. Available: https://www.itu.int/en/journal/j-fet/Pages/default.aspx.

[34] M. A. Aygül *et al.*, "Spectrum Occupancy Prediction Exploiting Time and Frequency Correlations Through 2D-LSTM," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020. [Online]. Available: DOI: 10.1109/VTC2020-Spring48590.2020.9129001.

[35] S. S. Shirgan and U. L. Bombale, "Hybrid Neural Network Based Wideband Spectrum Behavior Sensing Predictor for Cognitive Radio Application," *Sens Imaging*, Dec. 2020, doi: 10.1007/s11220-020-00293-4.

[36] B. G. Najashi and F. Wenjiang, "Cooperative Spectrum Occupancy based Spectrum Prediction Modeling," *Journal of Computational Information Systems*, 2014, doi: 10.12733/jcis10167.

[37] L. Yin, S. Yin, W. Hong, and S. Li, "Spectrum Behaviour Learning in Cognitive Radio based on Artificial Neural Network," in *MILCOM 2011 Military Communications Conference*, 2011. [Online]. Available: DOI: 10.1109/MILCOM.2011.6127671

[38] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature Selection in Machine Learning: A New Perspective," *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2017.11.077.

[39] B. B. H. D. Negash, "Techno-economic Analysis of LTE Deployment Scenarios for Emerging City: A Case of Adama, Ethiopia Addis Ababa, Ethiopia," in *International Conference on Information and Communication Technology for Development for Africa*, 2018. [Online]. Available: DOI:10.1007/978-3-030-26630-1_18

Image Deblurring with Compressive Sensing

Rahel Berhanu Dibeya[1,*], Fitsum Assamnew Andargie[1,*]

[1]*School of Electrical and Computer Engineering, College of Technology and Built Environment*
Addis Ababa University, Addis Ababa, Ethiopia
**Corresponding Author's E-mail address: mulrahel@yahoo.com, fitsum.assamnew@aait.edu.et*

DOI: https://doi.org/10.63990/zede.v43i.12974

ABSTRACT

Compressive sensing is a technique that enables recovery of signals represented by an underdetermined system of equations. Such a recovery of an original signal is made possible if the samples are represented in a sparse manner provided an appropriate measuring matrix is used for the modelled system. Blurred images are examples of signals that are sparse especially in transform domains. Different researches have been done to show the possibility of recovering blurred images that use sparse representation of transform domains by applying compressive sensing. In our work, however, we propose a model that doesn't require transforming into other domains. In addition, a box-wise approach has been used that derives the underdetermined system matrix from 7x7 segmented boxes of the blurred image. Compressive sensing algorithms are applied on these boxes to recover the whole image iteratively. Our method is shown to have a much better computational complexity than the traditional Lucy-Richardson deblurring method. Thus, with this improved computational complexity, the study provides an initial platform to deblur images using box-wise method and compressive sensing technique.

Keywords: Compressive Sensing, Deblurring, Gaussian blur, IHT, Sparsity.

1. INTRODUCTION

Image processing is a field which finds various applications in everyday life. In digital image processing, different operations are done one of which being image restoration. Image restoration is the process of recovering an original image from a noisy or blurred one [1]. One part of image restoration is image deblurring in which the blur is removed from a corrupted image by different techniques. In image deblurring, it is desired to have a restored image that is as much close to the original one by using minimal computational resources.

Presently, there exist different types of deblurring techniques. The Lucy-Richardson algorithm, neural network approach, deblurring with Noisy Image pairs and deblurring with Handling Outliers are few of them [2]. Among the different types of blurs encountered in image processing, Gaussian blur is one of them. For instance, Gaussian blur occurs in images taken in astronomy or medicine such as MRI.

Compressive sensing is a method which allows finding solution of equations which are underdetermined [3]. For reconstruction to be effective in compressive sensing, sparsity of the underdetermined system is one of the preconditions [4]. Blurred images are usually sparse in transform domains [5]. In this paper, a method which removes Gaussian blur using compressive sensing is introduced. With this method, a sparse system is derived and solved without transforming into other domains.

This research paper contains five parts. The first part is the introduction. In the second part, an overview of compressive sensing including definition, existing algorithms and related works are discussed. The third part details the modeling process used in this work. The fourth part includes the results obtained by applying the

compressive sensing algorithm and their comparison with other deblurring methods. Finally, conclusion and recommendation are given in the fifth part.

1.1 Compressive Sensing Overview

Compressive sensing is a theory first developed by David L. Donoho [3] and Candes, Romberg and Tao [6] in 2006 that can be used to solve a system of equations which is underdetermined. Starting from its first development, various algorithms have emerged that enable to solve an underdetermined system of equations. These algorithms have preconditions that need to be met for successful recovery of the solution. The first precondition is that the underdetermined system of equations should be represented in sparse manner [7]. The other common precondition for faithful recovery is incoherence which requires the matrix that represents the underdetermined system of equations to have low coherence [8]. The coherence of a matrix is defined as the absolute value of the maximum cross-correlations between its columns [9].

Let the underdetermined system of equations be represented by ϕ(*mxn*) which is obtained by multiplying some original set of equations, ψ(*nxn*), by a matrix M(*mxn*) known as a measuring matrix where *m*<<*n*. And let the data to be recovered be a vector X of length *n*. Then the system of equations which is underdetermined can be represented as [3]:

$$Y = \mathrm{M}\psi\mathrm{X}\,, \tag{1}$$

where ψ: *nxn* matrix

M: *mxn* matrix

X: *n* sized vector to be restored

Y: *m* sized known vector a.k.a observation vector

$$Y = \phi\mathrm{X} \tag{2}$$

where $\phi = \mathrm{M}\psi$: *mxn* matrix

For successful reconstruction to take place by compressive sensing, the matrix ϕ which is the product ψ and M should be sparse and have low coherence [7][8].

1.2 Restricted Isometry Property (RIP)

In compressive sensing, a parameter known as Restricted Isometry Property (RIP) must be satisfied to ensure the faithful recovery of signals. The importance of this parameter relies on the fact that low coherence of the ϕ matrix is closely related to it [8]. The Restricted Isometry Property of the matrix, ϕ, given by $\phi = \mathrm{M}\psi$, is defined as follows:

Assume a matrix $\phi_{mxn,}$ and an integer s such that $1 \leq s \leq n$. If a constant $\delta_s \in (0,1)$ exists such that, for every $m\ x\ s$ submatrix ϕ_s of ϕ, and for every vector z of dimension s [4],

$$(1 - \delta_s)\|z\|_2^2 \leq \|\phi_s z\|_2^2 \leq (1 + \delta_s)\|z\|_2^2 \tag{3}$$

holds true, then the matrix ϕ will have restricted isometry property of order s with the restricted isometric constant δ_s.

Most compressive sensing algorithms specify that the RIP value of the matrix ϕ to be less than a certain constant so that a successful reconstruction takes place. For example, Cande [8] specifies that the restricted isometric constant of order 2s should satisfy $\delta_{2s} < \sqrt{2} - 1$ for successful recovery in compressive sensing.

1.3 Sparsity

A vector or a matrix is said to be sparse if it consists of mainly zero elements. A vector or a matrix is said to be s-sparse if it has utmost s non-zero elements [10]. Sparsity enables to bring about efficient solutions in compressive sensing and different algorithms depend on it to recover a signal [7][11]. For a matrix $\phi_{mxn,}$ in (2), Donoho and Tanner [12] state that compressive sensing algorithms can recover most sparse signals if s is given by:

$$s = \frac{m}{2\log n} \tag{4}$$

1.4 Incoherence

In compressive sensing, it is important that the measurement matrix M is selected in a way such that it has the lowest possible coherence with ψ [9]. The mutual coherence, μ, for the matrix $\phi = M\psi$ which needs to be minimized is described by the expression given below [8].

$$\mu(\phi) = \frac{max}{i \neq j, 1 \leq i, j \leq n} \left\{ \frac{|\phi_i^T \phi_j|}{\|\phi_i\| . \|\phi_j\|} \right\} \quad (5)$$

where ϕ_i is the i[th] column of ϕ.

In the work by Tropp [13][14], it is stated that the coherence value, μ, need to satisfy

$$s < \frac{1}{2}(\mu^{-1} + 1) \quad (6)$$

for an accurate reconstruction to take place in the compressive sensing algorithms known as Orthogonal Matching Pursuit and Basic Pursuit.

1.5 Reconstruction Algorithms

In compressive sensing, there are different categories of reconstruction algorithms. The Matching Pursuit and Iterative Thresholding algorithms are two of them. The Matching Pursuit class of algorithms tries to represent a signal by linear expansion functions that form a dictionary [15]. Then the Matching Pursuit algorithm optimally selects dictionary elements that can best approximate the signal. The Orthogonal Matching Pursuit (OMP) [16] and Compressive Sampling Matching Pursuit (CoSaMP) [17] are examples of Matching Pursuit algorithms.

The Iterative Thresholding algorithms try to recover a signal iteratively. They use a thresholding function $\mathbf{H}s(\mathbf{x})$ at each iteration to set components of a vector X which are less than some number ε to zero and leave the rest of the components untouched [18]. Iterative thresholding algorithms include the Iterative Hard thresholding algorithms and Iterative Soft thresholding algorithms.

1.6 Related Works

Image denoising based on compressive sensing was done by A. Tavakoli and A. Pourmohammad[19] in which an additive noise was used to model the compressive sensing equation by $Y = \phi(X + Z)$ .The authors performed compressive sensing denoising using existing algorithms namely, Orthogonal Matching Pursuit(OMP) and Iterative Hard Thresholding (IHT) in which they illustrated that IHT is faster than OMP. They also compared compressive sensing denoising with classical filters like Wiener filter, Median filter, Wavelet denoising and Gaussian filters. And the results showed that compressive sensing denoising gave the same result as some of the classical filters or fairly better result than the rest of the existing methods.

In the work by Bruno Amizic et al. [5], blind image deconvolution is performed using compressive sensing. The authors experimented to show that a blurred image is mostly compressible in the transform domains. Based on this fact, the authors proposed a new algorithm that solves a constrained optimization problem. In doing so, they extended compressive sensing algorithms for use in blind image deconvolution and the experimental results from the work shows fairly better outputs than that of existing algorithms such as CoSamp.

Blind image deblurring using compressive sensing has also been performed by J. Yu et al. [20]. The work exploited the fact that similar structures usually recur in a natural image. The authors also exploited the fact that a natural image exhibits multiple similar patches or structures when the image is down sampled. Thus, the authors used the down sampled version of the blurred image in order to find sparse representation of the original image. Using structural multi-similarity and sparse representation a blind motion deblurring method was developed which was shown to have 98.88% success rate.

Metzler et al. [21] applied compressive sensing to already existing denoising methods. The authors integrated the existing denoising framework AMP (Approximate message passing) with compressive sensing recovery. In this method, they illustrated that using Denoising AMP and compressive sensing together gives state of the art recovery while operating ten times faster than existing denoising algorithms.

Compressive sensing image denoising is also done by Kang et al. [22]. In this work, the image was decomposed into edge and flat regions. In addition, an 8x8 measurement matrix was designed which was applied to the first three wavelet coefficients of the blurred image. Then from the existing compressive sensing algorithms, OMP (Orthogonal Matching Pursuit) was applied to construct each block in the image. Different error thresholds were used based on the block being in edge or flat region. Based on the experiments done by the authors, the proposed method gives better results than other existing methods.

2. MATERIALS AND METHODS

2.1 Applying Compressive Sensing to Deblur Gaussian Blur

Compressive sensing allows to solve an underdetermined system of equations given the sparsity and incoherence conditions are satisfied. It is known that there are different existing methods for sparse representation of signals. One of these methods commonly used for sparse representation is transformation of the blurred image into domains such as wavelet [5]. In this work, a non-blind deblurring that doesn't require transforming the blurred image into another domain has been done. The deblurring method is a non-blind one with a known kernel. The chosen kernel is a two-dimensional 7x7 Gaussian kernel with standard deviation of $\sigma = 2$.

2.2 Representing Gaussian Blur by a Sparse Matrix Using Convolution

In deriving a model for the deblurring, image convolution with the above-mentioned kernel is utilized. Convolution of an image involves replacing the pixels in an image with the linear combination of the neighbouring pixels according to the values in a certain kernel [23].

By applying the standard procedure of convolution between the stated Gaussian kernel and an image with width w and height h pixels, there will be a resulting wxh system of equations. Thus, given an image blurred with this kernel, if the original image is required to be restored, the following system of equations needs to be solved.

$$Y = \psi X, \quad (7)$$

where Y - the blurred image vector of length $s=wxh=wh$

ψ - an s x s matrix such that $s= wxh=wh$

X - the original image vector of length $s=wxh=wh$ to be restored

Basically, wxh is the size or the total number of pixels of the image. Therefore, there will be total number of equations which is equal to the total number of pixels in the image. To solve these equations in the direct way, the inverse of the matrix ψ has to be solved which usually has computational complexity of $O(s^3)$, where $s= wh$.

As mentioned earlier, the system of equations $Y = \psi X$ is obtained by convolving each pixel of the image with the convolution kernel. In these equations, a small fraction of the total number of pixels (7x7 or 49 neighboring pixels) is used to replace a pixel with their linear combinations. Because of this, the resulting equations are sparse mainly consisting of zeros. The sparsity found in these equations has made it possible to apply compressive sensing algorithms without transforming into other domains.

2.3 Specification of the Approach Used for Compressive Sensing Image Deblurring

While it is possible to apply compressive sensing algorithms to the whole set of equations which represents the total number of pixels, in this work however, an alternative method has been used. In this approach, the image has been segmented into 7x7 boxes measured in pixels. Then, 49 equations were derived resulting from convolution of every pixel in each box with the 7x7 Gaussian kernel. Out of the 49 equations, about half of them will be selected by a measurement matrix M. And to these selected equations, a compressive sensing algorithm is applied to retrieve the 49 original pixels. Finally, this procedure is repeated at each box iteratively until the whole image is covered.

The reason behind following this approach was initially to simplify the process of computing the RIP of the matrix to which compressive sensing was to be applied. In the end, however, the computation of the RIP parameter was not found necessary. This was because the Hard iterative Thresholding algorithm used in the work was found to converge because the norm-2 of the matrix involved was less than one.

Derivation of Sparse Basis Matrix ψ from a Gaussian Kernel

As the mentioned earlier, the method used for the deblurring involves segmentation of the image into 7x7 boxes. Thus, for an image with height h and width w pixel, the maximum value of the total number of boxes is given by:

$$Max.value\ of\ the\ total\ no.of\ boxes = (\frac{w}{7}+1\ )(\frac{h}{7}+1) \quad (8)$$

When a single 7x7 box is convolved with a 7X7 Gaussian kernel, there will be a resulting 49 set of equations with 49 unknown pixel values. These equations constitute the ψ matrix. Also, when a 7x7 box is convolved with a 7x7 Gaussian kernel only the central pixel at (3,3) is covered completely by the Gaussian kernel. Because of this, the convolution result will not be accurate at the rest of the pixels. Such effect is also observed when doing convolution at the edge of any image. For a convolution by a Gaussian kernel for example, this results in blackening of the image at the edges instead of whitening it. There are different existing methods that can be used to correct this such as wrapping or mirroring an image, ignoring edge pixels or duplicating edge pixels [23].

For the image convolution of 7x7 boxes at hand, ignoring the edge pixels cannot be an option because this will mean ignoring the whole image. But mirroring the image has been found a better option because symmetry is found frequently in nature and consequently in most pictures. After applying mirroring and doing the convolution of the 49 pixels of single box a normalized 49x49 ψ matrix will result.

2.4 Derivation of Measurement Matrix – M and Matrix – ϕ

The measurement matrix is *mxn* in size where $m<n$. The task of the measurement matrix is to reduce the n number of equations to m equations and it can be of different compositions. After doing the required computations and tests, finally the measurement matrix has been designed in such a way that it selects the even numbered rows from the main set of 49 equations. This measurement matrix is found to give the lowest coherence in the matrix ϕ which in turn gives favourable results in compressive sensing. Then, the ϕ matrix is obtained by using the relation

$$\phi = M\psi. \quad (9)$$

2.5 Selection of a Reconstruction Algorithm

There are different algorithms of compressive sensing that can be applied to a given problem. Here, the ϕ matrix which is derived from the image's convolution with the 2D-Gaussian kernel consists of coefficients that are all less than 1. As a

result, the norm-2 of the matrix becomes less than 1 which makes it suitable to apply the Iterative Hard Thresholding algorithm of compressive sensing. This is because the Iterative Hard Thresholding algorithm is stated to converge whenever the norm-2 of the matrix ϕ is less than 1 [18]. For the ϕ matrix that has been used here, the second norm is found to be: $\|\phi\|_2 = 0.7649$.

2.6 Computation of Parameters Required for Use in Iterative Hard Thresholding Algorithm

The Iterative Hard Thresholding algorithm is listed below [18].

Listing.1 The Hard Iterative Thresholding Algorithm

Input

- s the sparsity of X

-y ε R^m and the matrix ϕ ε R^{mxn}

Output:

- *X' such that y = ϕx'*

1. $X^{(0)} = 0$

2. for i=1,...do

3. $X^{(i)} = Hs(X^{(i-1)} + \phi^T(y - \phi X^{(i-1)}))$

4. end for

5. X'= $X^{(i)}$

where Hs is hard thresholding function which sets all components of the vector $\boldsymbol{x}$ *to zero except the s largest components.*

$$\boldsymbol{Hs}(\boldsymbol{x}) = \begin{cases} x_i, & x_i \geq \varepsilon \\ 0, & x_i < \varepsilon \end{cases},$$

ε is the s-largest component of X

Before applying the Iterative Hard Thresholding algorithm some parameters need to be predetermined.

Setting m where m is the number of rows of the matrix ϕ(mxn): The integer *m* is the number of rows of the matrix ϕ. For each 7x7 box of the image, 49 equations were derived making *n=49*. Compressive sensing uses underdetermined system of equations where *m<n*. In order minimize the computational complexity of the Iterative Hard Thresholding algorithm m should be as much small as possible. This is because the Iterative Hard Thresholding algorithm has computational complexity of O(*mn*) per iteration [24]. However, a balance should be kept in the minimizing of *m* which otherwise will compromise the restoration of the image. Keeping that in consideration, *m* is selected to be, *m = 16* for initial trial.

Setting the sparsity value s: According to Equation (4), *s* is calculated as:

$$s = \frac{m}{2\log n} = \frac{16}{2\log 49} = 4.733 \cong 5 \qquad (10)$$

Setting the upper bound ε of the thresholding function: *ε* is the s-largest element of X. It is also possible that *ε* can be selected randomly whenever there is no unique set of *s* number of elements which are the largest in the vector X [18]. The pixel values in an image range from 0-255. And most images contain much larger number of pixels than 255. As a result, most of the time there will not be found a unique set of *s* number of elements which are the largest pixel values in an image. Because of this, *ε* has been set randomly. Thus, *ε* is made to be equal to the sparsity value *s* giving *ε = 5*.

Computing theoretical maximum number of iterations k^*: The theoretical maximum number of iterations can be computed using Equations (11) and (12). And it holds true when the matrix ϕ has modified Restricted Isometric Property (RIP) of order 3s, $\beta_{3s} < 1/8$, where $\beta_s = 1 - \frac{1-\delta_s}{1+\delta_s}$. The calculation of RIP is a computationally intensive process and this is stated in [24]. Due to this, the Restricted Isometry Property (RIP), δ_s, has not been computed for the matrix ϕ . As it is mentioned before, the precondition for convergence of the Iterative Hard Thresholding algorithm is satisfied with the norm-2 of the matrix ϕ being less than 1. But with the RIP value of the matrix ϕ being unknown, the theoretical maximum

number of iterations may not hold true. As a result, the theoretical maximum number of iterations is computed and then taken as the starting number of iterations with which the Iterative Hard Thresholding algorithm is applied.

The theoretical maximum number of iterations, k^*, is computed as follows.

Given, $Y = \phi X + e$ (where e is the noise vector in the blurred image) and X' as the k^{th} approximation,

$$k^* = \left\lceil \log_2 \left(\frac{\|X'\|_2}{\varepsilon_s} \right) \right\rceil \quad (11)$$

$$\varepsilon_s = \|X - X'\|_2 + \frac{1}{\sqrt{s}} \|X - X'\|_1 + \|e\|_2 \quad (12)$$

$\|X'\|_2$ can have different values depending on the pixel color. Here, the maximum value of $\|X'\|_2$ is computed with the maximum value a pixel can have which is 255. X is n-dimensional with *n=49*. The norm-2 and norm-1 of the error vector, $\|X - X'\|$, is computed assuming a maximum of 1 pixel difference between the actual solution and the approximation vector. The noise is assumed to be zero giving $\|e\|_2 = 0$. With these values set, the theoretical maximum number of iterations becomes: $k^* = 6$

Initial approximation value: The initial approximation vector has been set to have a value of white pixel which is 255 i.e. $X^{(0)} = 255$.

3. RESULTS AND DISCUSSION

3.1 Specification of the Inputs Used for Deblurring

We used 24 bit bitmap images as inputs for testing our proposed work. Thus, the algorithm has been applied three times in each image for the R(red), G(green), B(blue) array of pixels.

The blurred images that are used as inputs in the deblurring process have been convolved box-wise. This means that when performing convolutions on the original images, the images are segmented into 7x7 pixel boxes and the convolution is done iteratively on each box separately. And the mirroring of pixels near the borders of each box has been done during convolution. This is not the natural way of performing convolution. It has been done in order to match the input blurred image with the model used for deblurring.

3.2 Applying IHT to Images and Results obtained

When IHT was applied to an image with the initial parameters computed in Sec. III an acceptable outcome was not obtained. In the resulting image, each box has not been restored sufficiently and because of that grids have been formed all over the image.

In order to improve the deblurring, m has been varied to different values and good results were obtained for *m*=25 by selecting even number of rows from ψ. This is because the ϕ matrix has the lowest correlation when such a selection is done. Table 1 gives few samples of correlation values of ϕ for different types of measuring matrix M where $\phi = M\psi$.

Table 1 Correlation values of different values of the measurement matrix

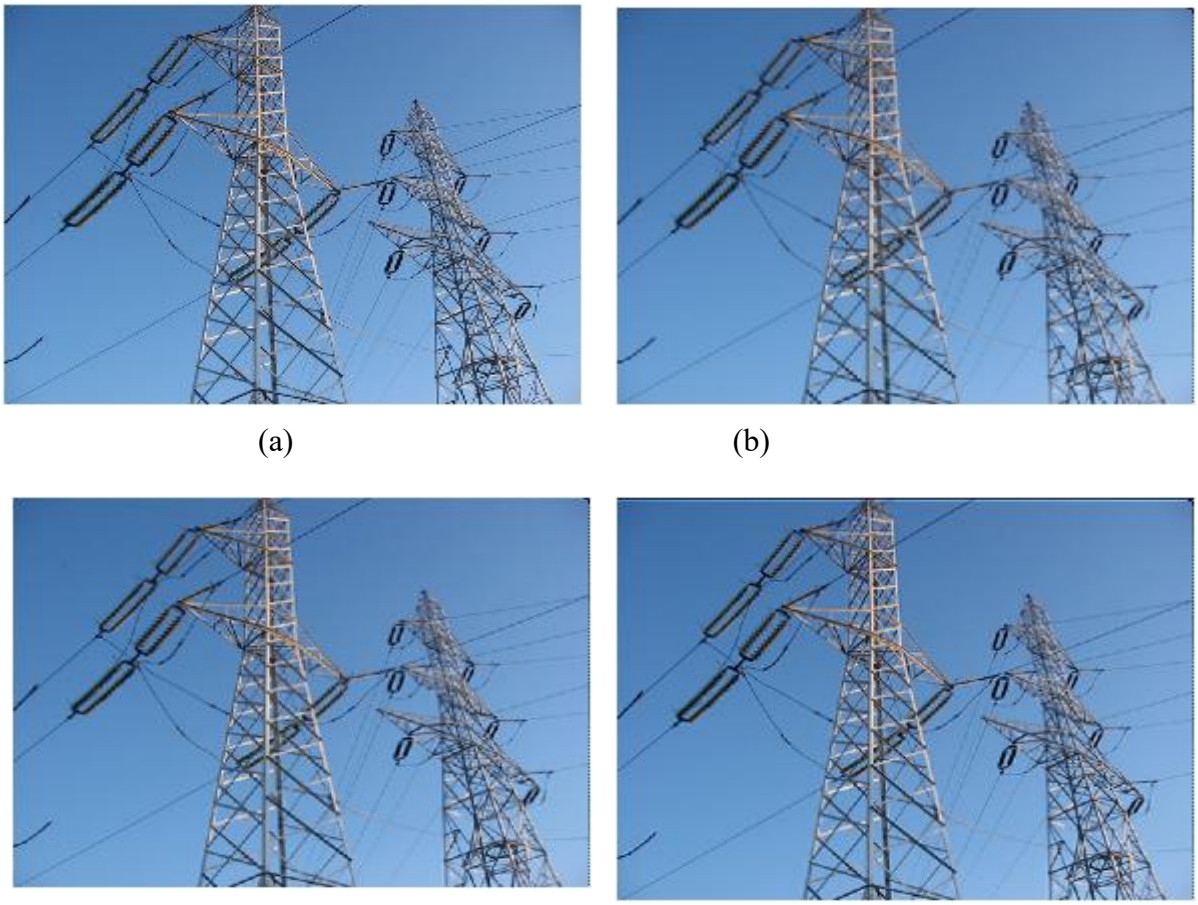
Type of Measurement matrix(M)	Value of *m*	Correlation of ϕ
Selects rows of ψ which are multiples of 3	16	0.979235
Selects rows of ψ which are even	25	0.953705
Selects rows of ψ randomly	32	0.975379

Iterative Hard Thresholding was applied to the convolved Electric_Lines image displayed in Figure 1 with the improved parameters shown in Table 2.

Table 2 Improved Parameters

No of rows of ϕ(m)	Upper bound of thresholding(ε)	No of Iteration(k)
25	5	6

In the deblurring result for the Electric_Lines image using *m=25* and the other parameters unchanged, an improvement of only 0.4% in Green pixels and 1 % in Blue pixels was obtained while the Red pixels deteriorated by 0.6%. Thus, the number of iterations was increased successively and the image was deblurred in better percentages. In the process, good results were obtained for $k = 45$. Results for $k = 12$ and $k = 45$ are shown in Figure 1c and Figure 1d.



(a) (b)

(c) (d)

Figure 1 (a) The original Electric_Lines image, (b) Box wise Gaussian convolved image, (c) Deblurring result for *m*=25, *k*=12, (d) Deblurring result for *m*=25, *k*=45

Results of IHT applied to a box-wise Gaussian convolved flower image are shown in Figures 2 (a to d).

(a) (b)

(c) (d)

Figure 2 (a) Original Flower Image, (b) Box-wise convolved Flower Image, (c) Deblurring result for k = 12, (d) Deblurring result for k = 45

3.3 Mean Squared Error (MSE) of the Deblurred Images

Mean Squared Error (MSE) of the deblurred images is shown in Table 3. The results show that IHT has deblurred the images to a measurable degree. In both images, the deblurring is negligible or negative at the theoretical maximum iteration *k=6*. As it has been stated earlier, this theoretical maximum iteration has been used as starting number of iterations since the RIP of the matrix ϕ is unknown. Thus, when the number iterations is increased to *k=12*, there is a considerable improvement in the deblurring ranging from 4%-30%. At iteration number *k=45*, the images show a much better deblurring result ranging from 10%-40%. During the test, it has been observed that the number of iterations cannot be increased from a certain upper bound beyond which the image will be degraded.

Table 3 MSE of the deblurred images

Image	MSE of Blurred image (RGB)	No. of Iterations-*k*	MSE of Deblurred image	Improvement
Electric_Lines	(R) 374.6 (G) 343.7 (B) 341.2	6	(R) 376.9	-0.6%
			(G) 342.0	0.4%
			(B) 337.9	1%
		12	(R) 349.5	6.7%
			(G) 316.0	8%
			(B) 308.8	9.4%
		45	(R) 317.2	15.5%
			(G) 289.1	15.9%
			(B) 272.0	20.3%
Flower	(R) 175.3 (G) 128.5 (B) 175.9	6	(R) 168.4	3.9%
			(G) 130.2	-1.3%
			(B) 169.7	3.5%
		12	(R) 123.5	30%
			(G) 113.5	11.7%
			(B) 133.0	22.4%
		45	(R) 105.2	40%
			(G) 115.3	10.3%
			(B) 120.7	31.4%

3.4 Computational Complexity of the Deblurring Method

The computational complexity analysis given below refers to both the time and storage requirement. The IHT algorithm has computational complexity of O(*mn*) per iteration [24]. For *k* number of iterations, the total computational complexity becomes O(*kmn*). The computational complexity of the deblurring method followed involves the cost of evaluating IHT at each box multiplied by the total number of boxes in the image. In addition, it includes the cost of evaluating ψ and ϕ matrices and the segmenting overhead.

For the analysis below,

n - represents the total number of pixels in a segmented box for this case a 7x7 box making *n=49*

m - represents the number of rows in the ∅ matrix making *m=25*

k - represents the number of iterations

Computational complexity of the evaluating the ψ matrix: Evaluating the ψ matrix includes representing the convolution of a 7x7 pixel box of the image with the 7x7 2D Gaussian kernel by 49x49 matrix. This process includes three steps. The first one is initial representation of the convolution coefficients and the 7x7 box pixel indexes by *nxn* matrix whose elements need to be ordered further. This step has computational complexity of $O(n^2)$. The next step is ordering of each row of the *nxn* matrix based on pixel indexes using a sorting algorithm. This step gives the initial ψ matirx before mirroring. The ordering or sorting algortihm used for this second step is insertion sorting. Insertion sort has computational complexity $O(n^2)$ and when it is applied for each of the *n* rows of the unordered ψ matrix, the resulting total computational complexity is $O(n^3)$**.** Finally, the ψ matrix is mirrored to account for the incomplete convolution near the edges of the box. This final step has computational complexity of $O(nxn) = O(n^2)$**.**

Thus, the total computational complexity of evaluating the ψ matrix is the sum of the complexities of the above processes and it is equivalent to $O(n^3)$.

Computational complexity of the evaluating the ∅ matrix: The ∅ matrix is given by $\emptyset = M\psi$ where M is *mxn* and ψ is *nxn*. This matrix multiplication has $O(mxnxn) = O(mn^2)$ computational complexity.

Computational complexity of applying IHT on a single 7x7 box: As stated above, Iterative Hard Thresholding, IHT, has computational complexity of *O(kmn).* In the box-wise approach, there is a cost of segmenting the image at each iteration which increases the computational complexity by a factor of 49. This is due to the cost of fetching an array of length 49 pixels from the total collection of the image's pixels at each iteration. Thus, the total computational complexity becomes *O(kmnSO)* where *SO = n = 49* is the segmenting overhead. Computational complexity of applying IHT a single 7x7 box $=O(kmnSO) = O(kmn^2)$

Computational complexity of applying IHT on the whole image: As it has been stated earlier, the total number of boxes in a 7x7 segmented image is given by:

$$Total\ no.of\ boxes = (\frac{w}{7} + 1)(\frac{h}{7} + 1) \cong \left(\frac{w}{7}\right)(\frac{h}{7})$$

Where *w* - width of the image

h - height of the image

Computational complexity of applying IHT on the whole image =

$O\left(kmn^2\left(\frac{w}{7}\right)\left(\frac{h}{7}\right)\right) = O(kmn^2\left(\frac{wh}{49}\right))$

$= O(kmn(wh))$

The total Computational complexity of deblurring the whole image: The total computational complexity of deblurring the whole image is approximately the largest of

the computational costs stated above. The largest computational cost is when applying IHT on the whole image. Thus, the total computational complexity of deblurring the whole image is $O(kmn(wh))$.

Computational Complexity of Applying IHT using non-Box-wise Method

If the Iterative Hard Thresholding algorithm had been used without segmenting the image into boxes, the expected computational complexity would have been as follows. In the case where non-box-wise method is used, the ψ matrix has dimensions *sxs* where *s = wh.*

For the analysis below

n - represents the total number of rows in the ψ matrix making *n = s = wh.*

m - represents the number of rows in the ∅ matrix making $m = \frac{n}{2} = \frac{wh}{2}$ assuming that half of the equations are selected.

k - represents the number of iterations as before.

Computational complexity of the evaluating the ψ matrix: Evaluating this computational complexity comprises the copying into, ordering and applying mirroring to the ψ matrix which has dimension *sxs* where *s=wh.* The computational complexity of these processes amounts to $O((wh)^3)$. This value is obtained by replacing the box-size with the image size, *n=s=wh,* in the previous derivation of the computational complexity of the ψ matrix in the box-wise deblurring method.

Computational complexity of the evaluating the ∅ matrix: As stated earlier, this computational complexity has order of multiplying the M*(mxn)* matrix by the $\psi(nxn)$ matrix to give the ∅*(mxn)* matrix. By replacing the *m* and *n* values are given above for the non-box wise deblurring case, this step will have a computational complexity of $O\left(\frac{(wh)^3}{2}\right) = O((wh)^3)$.

Computational complexity of applying IHT on the whole image: Substituting the values *k, n* and *m* for the non-box wise method stated above into the computational complexity of applying IHT to the box-wise method gives:

Computational complexity of IHT without segmenting into boxes = $O(knm)$=$O(k\frac{(wh)^2}{2}) = O(k(wh)^2)$

The total computational complexity of deblurring the whole image: The total computational complexity of deblurring the whole image by using non-box wise method is the largest computational complexity of the above steps. And this largest value is obtained when evaluating the ψ matrix giving computational complexity of $O((wh)^3)$.

3.5 Comparison of the Box-wise Method of Deblurring with Other Methods

As computed earlier, the box wise deblurring method has computational complexity of $O(kmn(wh))$. By substituting the values of *k*=45, *m*=25 and *n*=49, the computational complexity becomes $(55125(wh))$. A typical printing size of an image has a resolution of 540 x 360 pixels giving a total pixel size of *wh*=194,900[25]. Thus, for such an average sized image, it can be seen that the computational complexity is equivalent to $O((wh)^2)$. Applying IHT for the whole image, without using box-wise approach has computational complexity of $O((wh)^3)$. Directly inverting the ψ matrix by Gauss Jordan method to get the original image has computational complexity of $O((wh)^3)$ [26] while the Lucy-Richardson algorithm has computational complexity of $O(k(wh)^3)$ [27], [28].

3.6 Limitations of our work

When normally (non-box-wise) convolved images were attempted to be deblurred with the box-wise deblurring method, the process resulted in no deblurring at any number of iterations. The reason behind

this is the mirroring used in the 7x7 boxes of the images while deriving the deblurring model. Here, the box size and the kernel size used are the same resulting in only the central pixel to be convolved in the normal way without mirroring. This has resulted in too much approximation in the box-wise method so that a normally convolved image couldn't be deblurred using the given way.

4. CONCLUSIONS

In this paper, a box-wise method of deblurring images using compressive sensing has been introduced. When applied to box-wise convolved images, this box-wise deblurring method has been found to be computationally more efficient than using the non-box wise counterpart or using the direct matrix inversion method. The method also exhibited a better computational efficiency than the well-known Lucy-Richardson deblurring.

However, when the box-wise deblurring method was applied to normally convolved images the results were not desirable. As it is stated earlier, when the convolution is done box-wise, mirroring is applied at those pixel points which the kernel cannot cover completely. When the box size and the kernel size are the same, in the given case 7x7, only the central pixel of the box gets convolved in a normal way and for the rest of the 48 pixel mirroring must be done. That means, only 2% of the box is convolved in the conventional way which results in the mismatch of the deblurring model and normally convolved images.

To minimize the approximations resulting from mirroring, the box size can be increased. But this will be at the cost of increasing the computational complexity which is given by $O(kmn(wh)$ where n is the box size. Finding an optimum design which minimizes the number of iterations, k, and the value of m can compensate for the increase in computational complexity caused by increasing the box size. In doing these, the deblurring method could be applicable to normally convolved images without its computational complexity being compromised.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEGMENTS

The authors are deeply grateful to God the father, the Son and the Holy Spirit.

REFERENCES

[1] S. J. Reeves, "Image Restoration: Fundamentals of Image Restoration", in mic Press Library in Signal *Processing,* Elseviear, 2014, pp. 165-192.

[2] D. Singh, and . K. Sahu, R "A Survey on Various Image Deblurring Techniques", *International Journal of Advanced Research in computer and communication engineering,* vol. 2, no. 12, 2013.

[3] D. L. Donoho, "Compressive Sensing", I*EEE Transactions on Information Theory,* vol. 52, no. 4, pp. 1289-1306, 2006.

[4] E. J. Candes, and T. Tao, "Decoding by Linear Programming", *IEEE Transactions on Information Theory,* vol. 51, no. 12, pp. 4203-4215, 2005.

[5] B Amizic, L. Spinoulas, R. Molina, and A. K. Katsaggeolos, , "Compressive Blind Image Deconvolution", *IEEE Transactions on Image Processing,* vol. 22, no. 10, pp. 3994-4006, 2013.

[6] E. J. Candes, J. Romberg, and T. Tao, "Robust Uncertainty Principle: Exact Signal Reconstruction from Highly Incomplete Frequency Information", *IEEE Transactions on Information Theory,* vol. 52, no. 4, pp. 489-509, 2006.

[7] D. L. Donoho, "For Most Large Underdetermined Systems of Linear

Equations the Minimal 1-norm Solution is also the Sparsest Solution", vol. 59, no. 6, pp. 797-829, 2006.

[8] E. J. Candes, "The Restricted Isometry Property and its Implications for Compressed Sensing", 2008.

[9] J. A. Tropp, "Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise", vol. 52, no. 3, pp. 1030-1051, 2006.

[10] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to Compressive Sensing" [Online]. Available: webee.technion.ac.il.

[11] H. Cheng, "Modeling and Learning Visual Recongintion Theory Algorithms and Applications", 2015. [Online]. Available: www.Springer.com.

[12] D. L. Donoho, and J. Tanner, "Counting Faces of Random Projected Polytopes when the Projections Radically Lowers Dimensions", *Journal of the American Mathematical Society,* vol. 22, no. 1, 2006.

[13] J. A. Tropp, "Greed is Good: Algorithmic Results for Sparse Approximation", vol. 50, no. 10, pp. 2231-2242, 2004.

[14] J. A. Tropp, and A. C. Gilbert, "Signal Recovery from Random Measurement via Orthogonal Matching Pursuit", vol. 53, no. 12, 2007.

[15] S. G. Mallat, and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Transactions on Signal Processing,* 1993.

[16] S. G. Mallat, G. Davis, and Z. Zhang, "Adaptive Time Frequency Decompostions with Matching Pursuits", *SPIE Journal of Optical Engineering,* vol. 33, pp. 2183-2191, 1994.

[17] D. Needell, and J. A.Tropp, "CoSamp: Iterative signal recovery from incomplete and inaccurate samples", *Applied and Computational Harmonic Analysis,* vol. 26, no. 3, pp. 301-321, 2009.

[18] T. Blumensath, and M. E. Davis, "Iterative Thresholding for Sparse Approximations", *Journal of Fourier Analysis and Applications,* vol. 14, no. 5, pp. 629-654, 2008.

[19] A. Tavakoli, and A. Pourmohammad, "Image Denoising Based on Compressed Sensing", *International Journal of Computer Theory and Engineering,* vol. 4, no. 2, 2012.

[20] J. Yu, Z. Chang, C. Xiao and W. Sun, "Blind Image Deblurring Based on Sparse Representation and Structural Self-similarity", in 42th *International Conference on Accoustics,* Speech and Signal Processing, 2017.

[21] C. A. Metzler, A. Maleki, and R. Baraniuk, "From Denoising to Compressed Sensing", *IEEE Transactions on Information Theory,* 2014.

[22] W. Kang, E. Lee, S. Kim, D. Seo, and J. Paik, "Compressive Sensing Based Image Denoising Using Adaptive Multiple Samplings and Reconstruction Error Control", Proceedings of SPIE- The International Society for Optical Engineering, vol. 8365, 2012.

[23] Ludwig, J., *"Satellite Digital Image Analysis"*, [Online]. Available: web.pdx.edu.

[24] Blumensath, T. and Davis M. E., *"Iterative Hard Thresholding for Compressed Sensing"*, Applied and Computational Harmonic Analysis, vol. 27, no. 3, pp. 256-274, 2009.

[25] "*Image Resolution and DPI Requirement"*, Walgreen Co., 24 07 2017. [Online]. Available: http://wagcco.secure.force.com. [Accessed 28 09 2021].

[26] Farebrother, R., *"Linear Least Square Computations"*, STATISTICS, 1988.

[27] Lucy, L. B., "*An Iterative Technique for the Rectification of Observed Distributions"*, Astronomical Journal, vol. 79, no. 6, 1974.

[28] Richardson, W. H., "*Bayesian-Based Iterative Method of Image Restoration*", Journal of the Optical Society of America, vol. 62, no. 1, 1972.

Enhanced Mechanical Properties Characterization of General-Purpose Unsaturated Polyester Resin using Nano Cellulose Particles: Industrial Application

Angaw Chaklu [1]*, Araya Abera[1], Daniel Tilahun [1], Negash Getachew[2]

[1]*School of Mechanical and Industrial Engineering, College of Technology and Built Environment, Addis Ababa University, Addis Ababa, Ethiopia*
[2]*Department of Chemistry, Addis Ababa University, Addis Ababa, Ethiopia.*
[*]*Correspondence Author's E-mail address: Angawch1988@gmail.com*

DOI: https://doi.org/10.63990/zede.v43i.12975

ABSTRACT

Though the unsaturated polyester resin is highly versatile with a wide range of application, its low impact strength, low elongation at break and low toughness constrain its applications. In this article, study of an unsaturated polymer resin was conducted by adding optimum amount of nanocellulose fillers extracted from sugarcane bagasse. Mechanical and microstructural properties were characterized by testing the composite with 0.5 %, 1 %, 2 %, and 3 % fiber weight fraction of the nanocellulose. Mechanical properties such as tensile, compression, impact, flexural, as well as X-ray powder diffraction, Fourier Transform Infrared Spectroscopy (FTIR), and thermo gravimetric analysis (TGA) were performed. The results indicate that 10 nm particle size and 2 % nanocellulose by weight fraction is best, which gives enhanced mechanical properties of the composite material up to 45 % tensile strength, 38 % flexural strength, 13 % compression strength and 8 % impact strength improvement. The composite material also shows improved thermal stability and bond stretching due to the incorporation of nanocellulose particles.

Keywords: Enhanced mechanical property; Polyester resin; Nanocellulose; Toughness; Tensile strength.

1. INTRODUCTION

Polyester resins were first discovered by the Swedish chemist Berzelius in 1847 [1–3]. The unsaturated Polyester (USP) in particular is a versatile and low-cost condensation resin formed generally by the reaction of polyhydric alcohols with unsaturated dibasic acids [1]. Polymer matrix composites of cellulose nanofiller are widely used in various application areas like sanitary wares, furniture, water proofing, matrix for composite materials, water tanks, pipes, gratings and high-performance components. Agro-industrial wastes are increasingly recognized as valuable and cheaper bio-resources for development of renewable high-performance materials. Residue from annual crops such as wheat, rice, cotton, sisal, jute, and so forth, potentially provides an inexhaustible source of cellulosic fibers [4–5]. The marine and transportation industry (closure, body panels, fenders, boat construction etc.), aerospace, construction, shipbuilding, automobile industry are also the prominent users of cellulose nanofiller due to its excellent and rich hydroxyl group used for surface modification [6–10], though it shows low mechanical properties. Nanocellulose filler can be extracted from agricultural residues [11] such as sugarcane bagasse, cotton linter, and rice husks. Eco-friendly and green materials development paradigm approach aims to reduce the emissions of greenhouse gases to safeguard the environment at a global level and replacement of synthetically available

materials. The naturally occurring materials offer several advantages like sustainability, low cost, better resistance toward wear, low toxicity, high modulus of elasticity, and large specific area with better mechanical and biological properties [5].Nanocellulose filler can be used as a reinforcing material for composite materials which gives very high specific stiffness and strength [6]. Elastic modulus of nanocellulose is predicted more than the Kevlar fiber [10] and the specific strength is approximately 7–8 times higher than stainless steel [11]. Due to its low toxicity, renewable nature, nano scale dimension and good biocompatibility, the future scope of nanocellulose is very wide. Nanocellulose has rich hydroxyl groups which are used for surface treatment of materials. Mechanical properties of nanocellulose such as high stiffness and strength are attractive to use as filler material for composites which gives better mechanical and thermal properties. Nanocellulose composites are used in various applications such as in biomedical for tissue repair, drug delivery and implants of some body parts, in paper and packaging industry, in electronic industry such as time- temperature integrator, gas and leak detector, in structural material, sealant, etc. [12]. Maradini et al.[13] Presented the result for effect of using 0 – 5 % concentration nanocellulose filler (NCF) on curing, thermal and mechanical properties of epoxy polymers. The morphologies were examined using scanning electron microscopy (SEM) and transmission electron microscope (TEM), and the results show that NCF has positive effect on thermal, tensile, compression, flexural and water barrier properties. The overall results dictate that NCF has the potential as green nanofiller for epoxy polymer, but they did not study the effect of particle size, cost and environment. Omran et al. [14] also investigated how the micro and nanocellulose reinforcements yield enhanced mechanical properties of composite materials but did not investigate the possibility of using small particle sizes for more enhanced mechanical properties in industrial application using polyester. Embirsh et al. [15] presented the results of a study by using sugarcane nanocellulose and aluminum silicon carbide (Al-SiC) with polyester. The characterized results for tensile, compression, flexural, impact and thermal test gave enhanced mechanical properties compared with pure polyester resin for structural and other industrial applications, but they did not study the sugarcane nanocellulose and Al-SiC hybridization effect with polyester resin and didn't consider effect of particle size. Mateo et al. [8] reviewed the importance of agricultural wastes to produce nanocellulose particles which can be used as food packaging, mechanical reinforcement and water filtration process. The paper used acid hydrolysis and mechanical treatment methods for nanocellulose extraction which is costly and not safe for the environment. Nano materials have a typical grain size of less than 100 nm, whilst micro materials are characterized by grain size of less than 500 nm up to 100 nm [1, 15, 17], but it is challenging to extract nanoparticles to the required particle size.

Unsaturated polyester resin has low mechanical properties and is not preferred in the industrial and structural [16, 18] applications due to its low toughness, low impact strength, low elongation at break and poor resistance to crack propagation [1, 19], which limits its application as weak polymer matrix material. So, there is a need to enhance its toughness and strength by reinforcing using nanocellulose fillers. Based on the literature, mechanical properties of composite materials can be modified using potential substrate materials and nanoparticles [20–21]. In general, mechanical properties enhancement of composite materials can be achieved using filler size variation, hybrid bonding of nanocellulose fillers or use of alternative potential green material as filler has not been addressed well and needs further investigation. This research focuses on use

of potential nanocellulose green material with nanocellulose filler size of 10 nm from sugar cane bagasse ash (SCBA) at Wenji Showa sugar factory. The aim is to enhance mechanical properties such as strength and toughness of unsaturated polyester resin by incorporating nanocellulose fillers for industrial applications.

2. MATERIALS AND METHODS

2.1. Matrix- Unsaturated Polyester Resin

Unsaturated polyester resin was prepared by esterification reaction of unsaturated dibasic acids with dihydric alcohols [1]. It is mostly used as polymer matrix material due to its excellent adhesion, low cure shrinkage, low cost and availability with a proportional combination of resin to hardener in the ratio of 10:1 as recommended in [22–24].

Hardeners are used as a catalyst to cure the resin properly by making a chemical reaction without changing its own composition and they facilitate solidification process of unsaturated polyester resin and filler materials from liquid to solid state. The curing agent used for this research was methyl ethyl ketone peroxide [MEKP] due to its good mechanical curing capability, availability and because it does not affect the mechanical properties of the individual composite materials [1]. The physical and mechanical properties of general purpose commercial unsaturated polyester resin used for this study were density 1.103 g/cm^3 [12], dynamic viscosity 11.789 Pa.s [12], specific gravity 1.2 g/cc [12], tensile strength 28.25–78.85 MPa [12], tensile and flexural modulus 1978.5–3000 MPa [25], flexural strength 44.50–120 MPa[25], compression strength 110.00-150 MPa [1, 12, 25], volume shrinkage 7% [25], tensile elongation 2.5% [25], fracture energy 90 J/m^2 [17, 25], impact energy at 25 ºC 3.5–6.5 Joule [17, 25], hardness 31.5–48 BHN [25], fracture toughness 0.30 MPam$^{0.5}$ [25] and Young's modulus 1±0.4 GPa [25–31].

2.2. Fiber-Nanocellulose Particle Extraction

Sugarcane bagasse was collected from Wenji Showa sugar factory and washed thoroughly with distilled water. Later, it was boiled in a vertical autoclave for one hour at 110 ºC to remove water-soluble materials. It was then sun-dried to remove the absorbed moisture. Later, the dried sugarcane bagasse was crushed and soaked in a 15 wt. % NaOH solution at room temperature to remove water insoluble materials such as lignin, pectin, natural oil, and wax that shields the surface and alkaline from the bagasse as illustrated in the Figure 1. Following that, the bagasse was exposed for drying. The alkali-treated fibers were then soaked in a 1 mol HCl solution at 70 ºC to dissolve the cell walls and detach the micro fibrils. To remove the remaining nanocellulose components, the treated bagasse was re-soaked in a 2 wt. % NaOH solution at 70 °C after the fibers had been thoroughly washed with deionized water to remove the hydrochloric acid. The fibers were cleaned and sun dried after being treated. To produce the nano cellulose from sugarcane bagasse, these fibers were finely ground at a high-speed grinder for 20 min at a revolution of 1500 rpm [4]. The importance of having fine nanocellulose particles such as 10 nm has great ability to bridging effect of polyester resin so that its strength can be enhanced well. Figures 1 and 2 summarize the nanocellulose extraction process.

2.3. Composite Material Preparation

150 mL resin was poured into a big separate bowl and nanocellulose was added to the bowl in the ratio of 0.5%, 1%, 2%, 3% by weight fraction of the total bowl volume for one-time sample preparation and the mixture was poured into the standard mold. The mixture was homogenized by stirring vigorously at 660 rpm for three minutes to have good nanocellulose dispersion as shown by SEM image in Fig.8. The hardener methyl ethyl ketone peroxide [MEKP] was added gradually, with the type

chosen depending on the mold material type and the desired properties of the final parts. Paste wax and polyethylene plastic were used as mold releasing agent to enhance the surface quality of composite and safe removal of the sample. Curing of the sample was done in open space environmental condition until the sample got good strength to remove from the mold.

(a) (b) (c) (d)

Figure 1 Nanocellulose extraction process: (a) Sugarcane bagasse, (b) Crusher, (c) Crushed bagasse, (d) Crushed cellulose

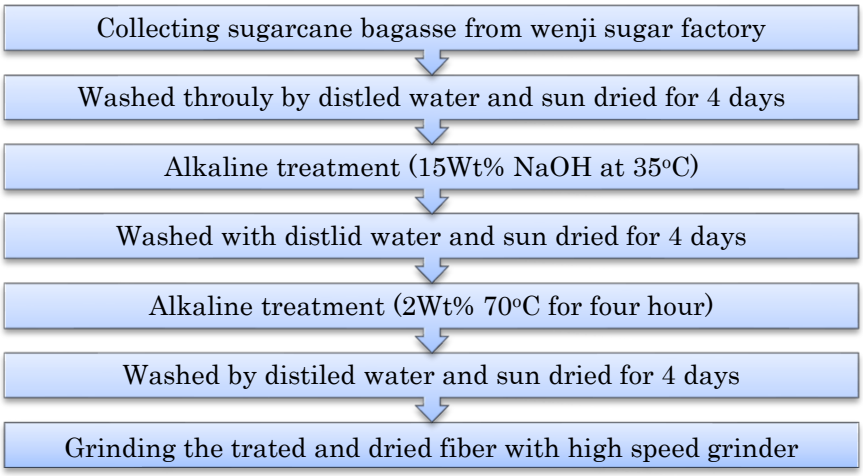


Figure 2 Nanocellulose extraction from sugarcane bagasse collection and removal process.

2.4. Characterization Setup and Apparatus for the Composite Materials

This research was carried out using experimental method via polymer composite material test standard to compare with literature value for validation. A computer-controlled small punch creep testing machine that can handle loads up to 200 kN was used during the test. The machine comes with a load cell capable of measuring up to 200 kN. The machine was used to perform a series of tests, such as assessing tensile strength according to ASTM D638-14 standards, as well as conducting compression strength and high-temperature tensile tests [21]. The miniflex 600 powder XRD machine for particle size determination, HCT_1 compressive thermal analyzer was considered for the measurement of thermal characteristics of the composite material and measuring device thermo scientific Nicolet is 50 FTIR with a range of (12000 – 50 cm^{-1}) and resolution of 0.125 cm^{-1} was used for the spectroscopy testing process.

2.4.1. Tensile test Setup and Apparatus

For the tensile test of this polyester resin with nano cellulose composite material, proportional additives of 0.5%, 1%, 2% and 3% (by weight fraction of fiber to matrix) were used. Samples were prepared and tested using polymer ASTM standard as

shown in the Figure 3 with ambient temperature of 23±2 ^{0}C and humidity value of 50±5 % as per ASTM D638-14 statement.

$$\sigma_t = \frac{P}{bh} \quad [25–30] \qquad (1)$$

Where, P= Failure load (N), B= Specimen's width (mm), H= Specimen's thickness (mm), σ_t=Tensile Strength (MPa).

Figure 3 Tensile testing machine setup

2.4.2. Flexural Characterization test Setup and Apparatus

Flexural mechanical property is also another point of interest to be enhanced using the ASTM test standards via UTM machine with capacity of 100 kN, load cell 300 kN, strain rate 5mm/min at ambient temperature of 25^{o}C with relative humidity of 62% as per ASTM D-790+17 for unsaturated polyester resin and Nano cellulose additive particles as a composite material shown in Figure 4.

$$\sigma_f = \frac{3PL}{bh^2} \quad [25–30] \qquad (2)$$

Where, P= maximum load (N), L=Specimen's Span length (mm), b= Specimen's width (mm), h= Specimen's thickness (mm), σ_f=Flexural Strength (MPa).

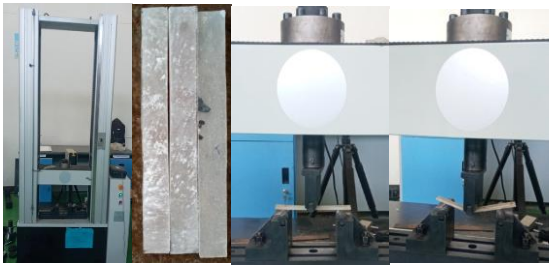


Figure 4 Three point flexural testing UTM machine setup

2.4.3. Impact Characterization Setup and Apparatus

Notched Charpy impact test using Charpy Izoid Impact testing machine shown in Figure 5 with capacity of 90J was carried out. In each type, specimens were tested in the ambient condition according to the ASTM D-256 standards with ambient temperature of 25^o C and approximate humidity value of 50±10% as per ASTM D-256. The average value was noted at the initial load deformation and tabulated as impact strength.

$$IS = \frac{A}{bxh} \quad [25–30] \qquad (3)$$

Where; IS=Impact strength, A=Energy consumed by Impact Specimen (Jm), b=width (mm) of specimen from middle of the notch, h=thickness (mm) of specimen from middle of the notch.

Figure 5 Charpy Izoid Impact testing machine setup

2.4.4. Compression Characterization test setup and Apparatus

Compression mechanical property analysis was done using the WDM-100S universal testing machine (UTM) shown in Figure 6 with loading velocity of 2mm/min as a strain rate, load cell 300 kN and test samples were done at temperature of 23±2 ^{o}C and humidity of 55±5% as per ASTM D-695-10 standards description [27].

$$\sigma_C = \frac{P}{bh} \quad [25–30] \qquad (4)$$

Where p=failure load (N), b=Specimen's width (mm)h= Specimen's thickness (mm), $\sigma_{C=}$Compression strength (MPa).

Figure 6 Compression testing machine setup

3. RESULTS AND DISCUSSION

3.1. X-Ray powder Diffraction of Nanocellulose

A miniflex 600 powder XRD, Tokyo, Japan machine was used. A micro focus X-ray tube with a Cu kα target was used as X-ray source, producing radiation (λ=1.5406A^0) at 40kV and 15 mA In the 2θ range of 10^0-70^0 with a step size of 0.01^0 on pure cellulose powder to determine crystallite size using Debye-Scherrer [1, 28, 32] equation. The raw XRD data were smoothed using origin 2018 software to obtain the full width at half maximum (FWHM) and peak positions of the particle as shown in Figure7. The figure shows the different peaks analyzed with baseline using peak analyzer features of the software and an average nanofiller crystallite size of 10 nm is obtained. FWHM (β) value is 0.945981 radians with peak position (2θ) is 58.46614^0 & λ=0.15406nm has been used.

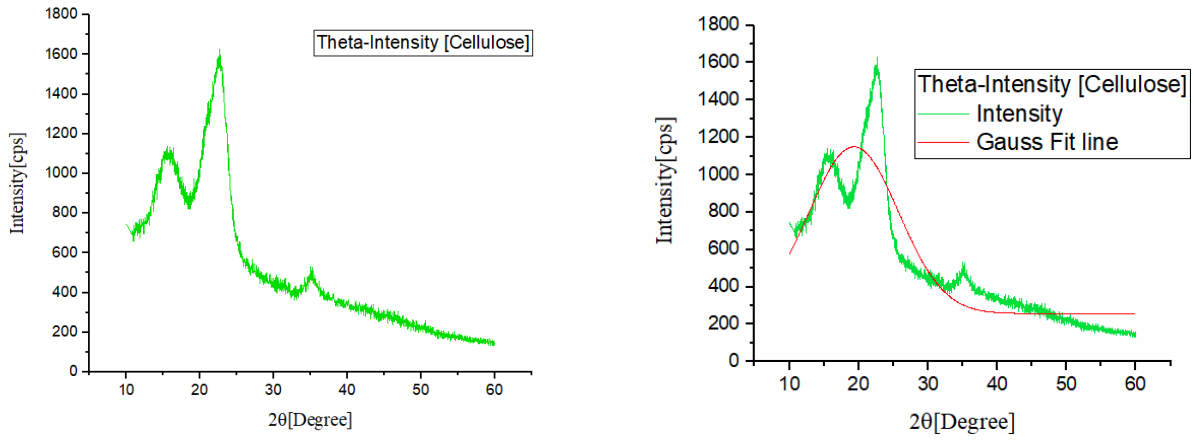
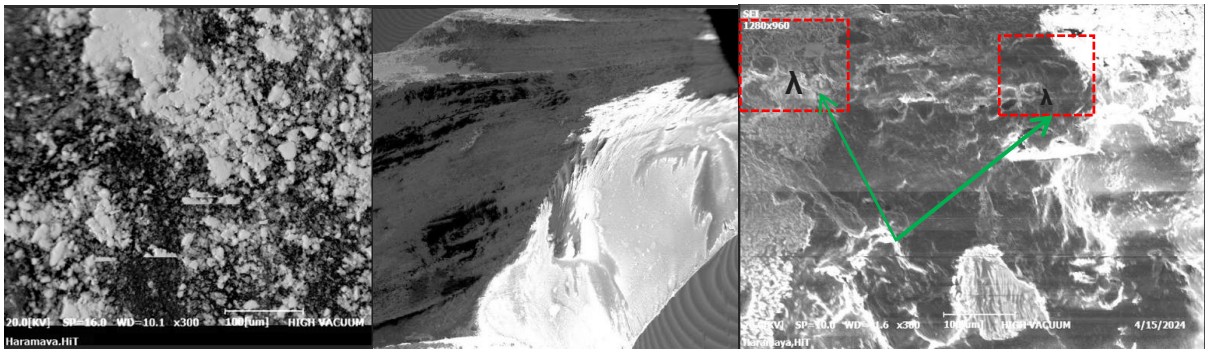


Figure 7 The XRD result after performing baseline correction & peak analysis

3.2. SEM Image of Nanocellulose



(a) (b) (c)

Figure 8 SEM image of **a)** Pure cellulose with a scale bar of 100 µm and resolution 300X, **b)** Pure rsin with a scale bar of 100 µm and resolution 300X, c) Composite cellulose(2%) and resin with a scale bar of 100 µm and resolution 300X

3.3. Mechanical Test Results of the Composite Material

3.3.1. Tensile Test

The stress – strain plots of the tensile test results are summarized in Figure 9.

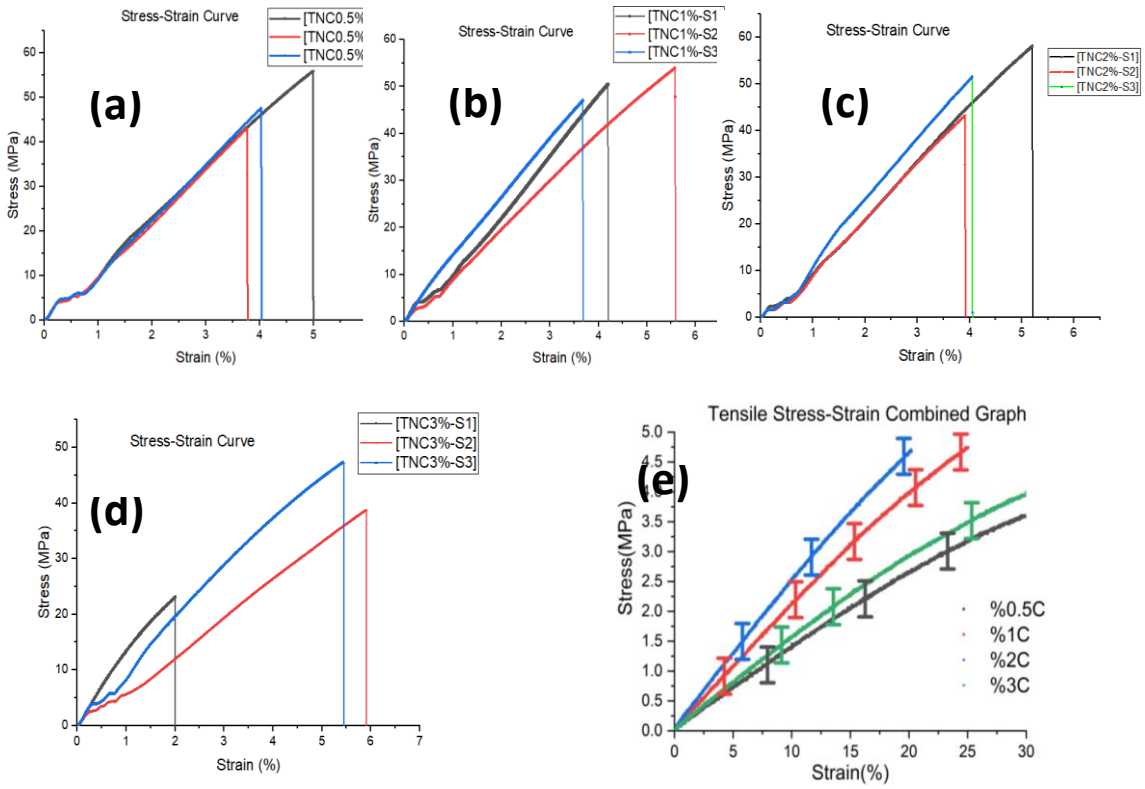


Figure 9 Plots of tensile test results: (a) 0.5 % nanocellulose and polyester resin, (b) 1 % nanocellulose and polyester resin, (c) 2 % nanocellulose and polyester resin, (d) 3 % nanocellulose and polyester resin and (e) (0,5%,1%,2% & 3 %) nanocellulose and polyester resin combined.

3.3.2. Flexural Test

The flexural test results are summarized in Figure10.

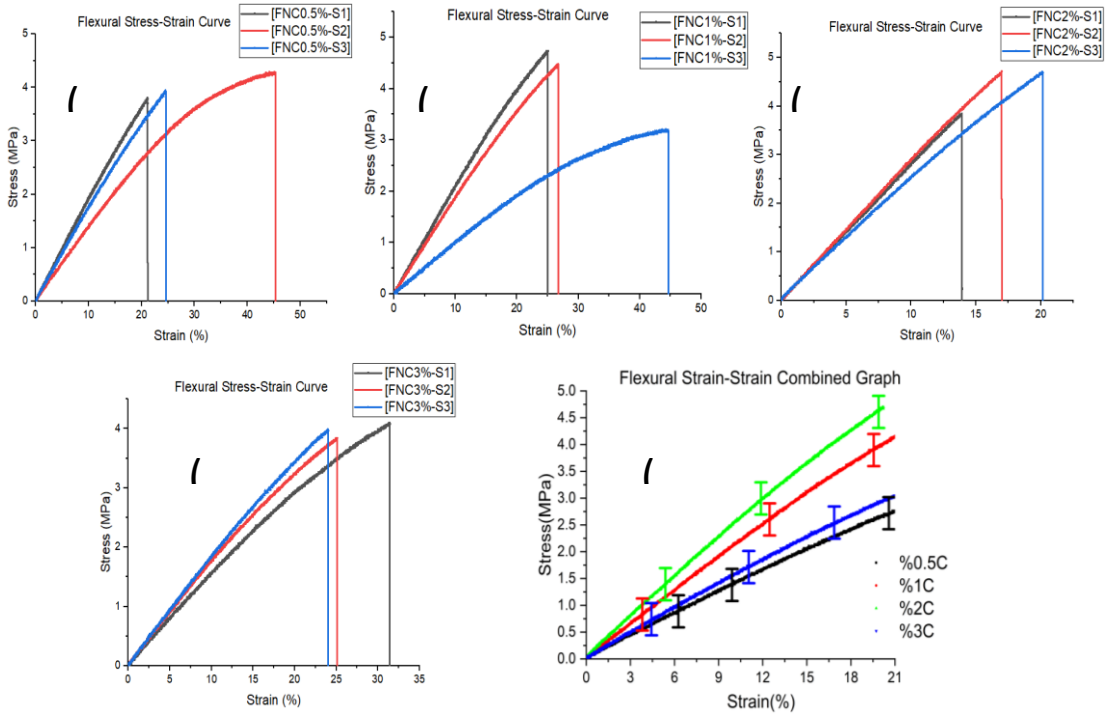


Figure 10 Plots of the flexural test result (a) 0.5 % nanocellulose and polyester resin, (b) 1 % nanocellulose and polyester resin, (c) 2 % nanocellulose and polyester resin, (d) 3 % nanocellulose and polyester resin and (e) (0,5%,1%,2% & 3 %) nanocellulose and polyester resin combined graph shown below.

Table1 Tensile and Flexural tests

Tensile test Sample code	Stress (MPa	Mean Dev.(MPa)	standard dev.(MPa)	Flexural test Sample code	Failurestress(MPa)	Mean Dev.(MPa)	Standard Dev.(MPa)
TNC0.5 %-S1	53	46.5	46.5±9.19	FNC0.5%-S1	152.40	116.20	116.20±51.1
TNC0.5 %-S2	43	41.5	41.5±2.12	FNC0.5%-S2	176.46	128.23	128.23±68.2
TNC0.5 %-S3	48	44.0	44.0±5.66	FNC0.5%-S3	160.42	120.21	120.21±56.8
TNC1 %-S4	51	45.5	45.5±7.78	FNC1%-S4	192.51	136.25	136.25±79.5
TNC1 %-S5	54	47.0	47.0±9.90	FNC1%-S5	184.48	132.24	132.24±73.8
TNC1 %-S6	47	43.5	43.5±4.95	FNC1%-S6	128.34	104.17	104.17±34.1
TNC 2 %-S7	58	49.0	49.0±12.7	FNC2%-S7	160.42	120.21	120.21±56.8
TNC 2 %-S8	43	41.5	41.5±2.12	FNC2%-S8	192.51	136.25	136.25±79.5
TNC 2 %-S9	52	46.0	46.0±8.49	FNC2%-S9	192.51	136.25	136.25±79.5
TNC 3 %-S10	23	31.5	31.5±12.0	FNC3%-S10	168.44	124.24	124.24±62.5
TNC 3 %-S11	39	39.5	39.5±0.71	FNC3%-S11	152.40	116.20	116.20±51.1
TNC 3 %-S12	47	43.5	43.5±4.95	FNC3%-S12	160.42	120.21	120.21±56.8

3.3.3 Impact Test

The mixture of unsaturated polyester resin and cellulose fillers (USP + cellulose filler) shows a slightly higher impact resistance than pure unsaturated polyester resin, according to reports in the literature [1, 3]. The increase in impact strength is 8%.

When conducting the test, the machine determines the average value for different amounts of cellulose filler [25].

3.3.4 Compression Test

In the first sample batch, the compression test results are as shown in Table 2. For CNC0.5 %-S1 (compression test of nanocellulose 0.5% for composite sample batch-1) failed under a failure load of 33.15 kN, CNC1 %-S1 failed under a maximum failure load of 22.846 kN, CNC2 %-S1 failed at a maximum failure load of 28.262 kN and finally CNC3 %-S2 failed at a maximum failure load of 26.339 kN.

Table 2 Compression and Impact test

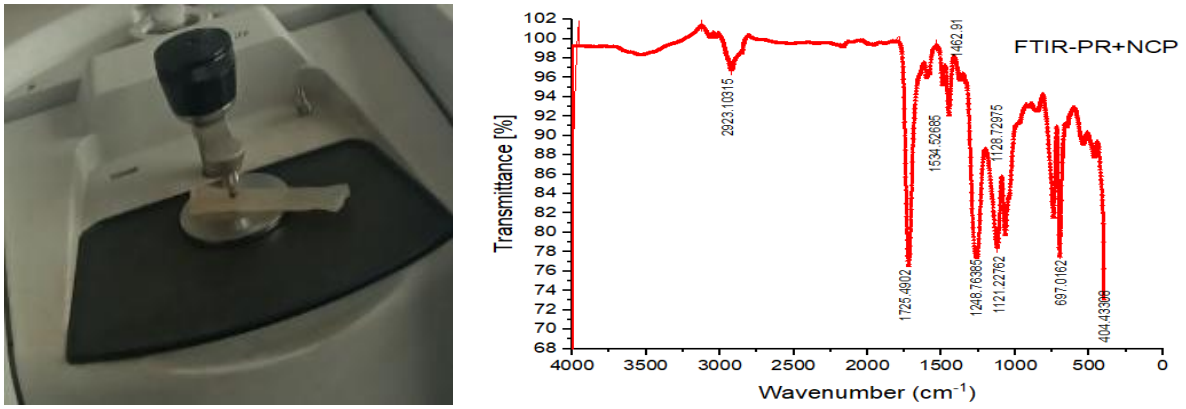
Compression test Sample code	Compression Stress (MPa)	Mean Deviation (MPa)	standard deviation (MPa)	Impact test Sample code	Failure Energy (J)	Mean Dev. (MPa)	Standard deviation (MPa)
CNC0.5 %-S1	114.72	112.36	112.36±3.34	INC0.5%-S1	4.4	4.4	4.45±0.14
CNC0.5 %-S2	132.60	121.30	121.30±15.9	INC0.5%-S2	4.5	4.5	4.50±0.00
CNC0.5 %-S3	84.60	97.30	97.30±17.96	INC0.5%-S3	4.3	4.4	4.40±0.28
CNC1 %-S4	91.40	100.70	100.7±13.15	INC1%-S4	4.4	4.45	4.45±0.14
CNC1 %-S5	67.08	88.540	88.54±30.35	INC1%-S5	4.3	4.4	4.40±0.28
CNC2%-S6	113.04	111.52	111.52±2.15	INC1%-S6	4.2	4.35	4.35±0.42
CNC2%-S7	118.40	110.70	110.7±0.99	INC2%-S7	4.6	4.55	4.55±0.14
CNC2%-S8	123.00	112.00	112.0±2.83	INC2%-S8	4.4	4.45	4.45±0.14
CNC3%-S9	107.20	105.60	105.6±6.22	INC2%-S9	4.5	4.5	4.50±0.00
CNC3%-S10	105.36	107.68	107.68±3.28	INC3%-S10	4.2	4.35	4.35±0.42
CNC3%-S11	78.20	94.10	94.10±22.48	INC3%-S11	4.5	4.5	4.50±0.00

3.3.5. Fourier Transform Infrared Spectroscopy (FTIR) Analysis

Thermo Nicolet iS50 Fourier Transform Infrared Spectroscopy (FTIR) Thermo Fisher Scientific, Waltham spectroscopy was used for the study. Samples were placed on sample holder after a carful cleaning with acetone solution. The machine can use either solid, liquid or powder sample types. Thermo Nicolet iS50 instrument uses the range of 400–4500 cm^{-1} to analyze stretching frequency of nanocellulose composite.

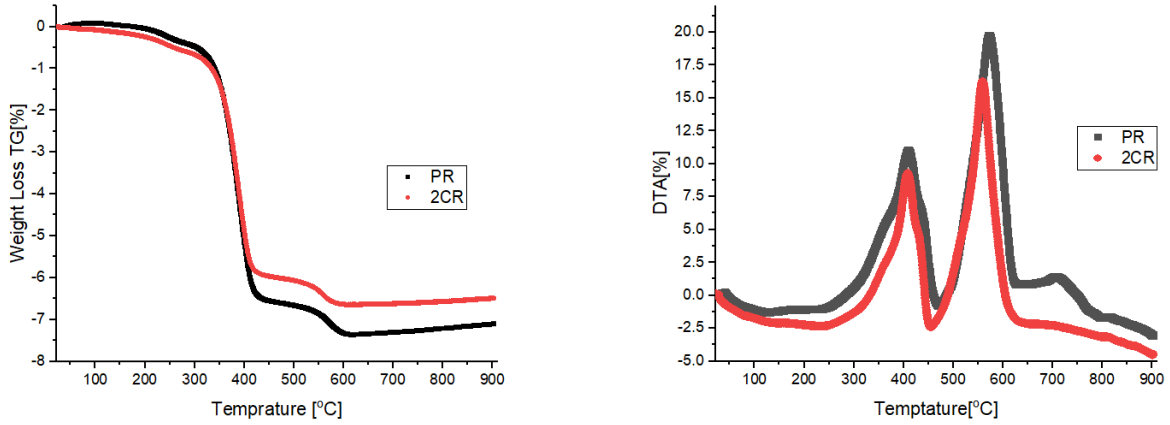
3.3.6 Thermo-Gravimetric Analysis of Nanocellulose with Resin

TSA 409 PC Luxx simultaneous thermal analyzer was used to determine nanocellulose composites' thermal stability. First, powder form of the composite was prepared, then 12 mg of powder was loaded into TGA crucible and heated at 10 ^{0}C/min in nitrogen atmosphere, in the temperature range of 25^{0}C to 900^{0}C as shown in Figure 12. TGA analysis of the composite material and pure polyester resin material were done and compared at maximum weight loss (Tdmax), which is more effective than five percent weight loss (Td5%) and ten percent weight loss (Td10%) thermal stability parameters.



(a) (b)

Figure 11 FTIR results, a) FTIR testing machine; b) spectra result of 2 % cellulose Nanoparticles particles (NCP) with pure resin (PR) composite material



(a) (b)

Figure12 TGA results (a) pure resin (PR) and 2 % cellulose resin (CR) composite vs temperature and (b) pure resin (PR) and 2 % cellulose resin (CR) vs Temperature differential thermal analysis (DTA)

3.4. Discussion

Tensile test in Figure 9 indicates maximum tensile stress for 2%C compared to 0.5%C, 1%C, and 3%C and this is in good agreement with previous literature [1, 4, 11]. Non-uniform dispersion of the nanocellulose particle may lead to particle agglomeration which can reduce material tensile strength. This is avoided during composite manufacturing phase by considering proper filler homogenization and more sample preparation to substitute premature failure. The composite material enhances its tensile strength about 45% upon the addition of nanocellulose filler in comparison with pure polyester resin tensile strength of 40MPa obtained from preliminary testing of pure polyester resin as well as comparing with literature value [1, 9]. Mean deviation is a measure of dispersion of tensile test data deviation from preliminary as well as literature value implying in good agreement with this study as shown in Table 1 and Fig.9. Flexural test in Figure 10, indicates maximum bending strength is obtained at 2%C compared with 0.5%C, 1%C, and 3%C nanocellulose by weight composition [2,9] implying that addition of nanocellulose filler provides additional strength to the composite material. Flexural strength of the composite is enhanced about 38% up on the addition of nanocellulose filler compared with polyester resin preliminary and literature

based flexural strength of 80MPa [4, 9]. Compressive test in results provided in Table 2 indicates improved value at 2%C compared with 0.5%C, 1%C and 3%C values. The composite material develops enhanced compressive strength about 13% compared with pure polyester resin preliminary accepted compressive strength of 110MPa [1, 9]. Its deviation is under the expected error range and in good agreement with literature [1]. As the amount of cellulose particle is more added, the composite brittleness increases which leads to catastrophic failure. Higher percentage cellulose filler contents show diminished mechanical properties of the composite material due to void formation, inability to fill gap between molecules, made the polyester highly viscose which imply formation of agglomeration and hinders uniform dispersion. For impact test as shown in Table 2, an eight (8%) enhancement has been achieved compared with polyester resin preliminary value of 4.5J [9] with smooth standard deviation. As the percentage of the nanocellulose increase, ability to absorb energy enhanced for this characterization but with slow percentage.

Generally, incorporating nanocellulose into polyester resin enhances interfacial bonding implying mechanical property improvement of the composite at 2%C by weight [1] for tensile, flexural, compression and impact test.

SEM image in Figure 8 indicates individual particle size and homogenously dispersed composite material structure resulting increased interfacial adhesion due to uniform dispersion of nanocellulose within the matrix which enhances effective interaction between nanocellulose and resin fillers that imply improved tensile strength. Some porosity sign of the composite material is observed which imply careful homogenization is needed during sample preparation and this may lead to failure of composite material [1, 27].

The chemical constitutes of nanocellulose composite and pure nanocellulose was evaluated as shown in Figure 11. The peak shift due to presence of various functional groups between 6000 cm^{-1} and 500 cm^{-1} has been observed. The peak observed for composite material at 1716.71 cm^{-1} corresponds to stretching of the C-O vibration group. The peak observed at 2923.1 cm^{-1} is attributed to O-H stretching vibration group of hydroxyl functional group of the composite material [1, 3]. The observed peak at 1644.47 cm^{-1} confirmed the presence of C-O bonds in polysaccharide aromatic rings of cellulose providing valuable information about the resin's unsaturation level and potential for cross-linking and the peak at 2923.1 cm^{-1} indicated the presence of single C-H bond in cellulose with C-H bending group [1, 9]. In contrast with pure cellulose particle, composite cellulose material with resin has higher peak value and new peak value has been created implying the presence of increased hydrogen bonds due to homogenized mixing of unsaturated polyester resin identified via tensile testing [1,4]. Figure 12a shows that initially, both composite materials lost some weight due to evaporation of moisture but its rate of degradation increased between 200^0C to 300^0C. However, increasing the temperature between 35^0C and 400^0C drastically increased weight loss of the composites. Particularly unsaturated polyester resin weight loss increased compared to nanocellulose composite. This is due to limited thermal conductivity of polyester resin [1, 4]. However, the addition of nanocellulose particle to the resin enhances thermal stability by making uniform distribution of heat through the composite which is enabled by uniform dispersion of nanocellulose particles than localized regions which cause high thermal weight degradation [4]. Uniform particle dispersion allows for efficient heat transfer within the composite by reducing localized hot spot and thermal stability of the composite is improved.

4. CONCLUSIONS

This research reports the experimental investigation to enhance mechanical properties of the composite unsaturated polyester resin material. The study has shown that the tensile strength, flexural strength, compression strength and impact strength have improved up on the addition of 2 % by weight of nanocellulose filler. The enhanced composite material can be used for water tank construction, water proofing work, matrix for composite materials and gratings.

Future research direction can rely on expanding the polyester resin material for structural application using hybrid bonding of nanofiller with cellulose, silica and rubber for more enhanced mechanical properties with wider scope of application. Furthermore, continuation of this study can include thermal stability at different levels of weight loss, dynamic mechanical thermal analysis (DMTA). Addressing the practical transition from laboratory scale to commercial mass production in order to achieve final materials which will be introduced in the market especially optimization of the whole process and development of new methods to produce Nano cellulose-based materials as well as reducing high production cost can be considered in the future.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

The authors thank Addis Ababa Science and Technology University (AASTU) for its support.

REFERENCES

[1]. E. Gashawtena, A. Kidane, and B. Sirahbizu, (2024). “The effect of nanocellulose and silica filler on the mechanical properties of natural fiber polymer matrix composites.” *Results in Engineering*, *24*, 102898.

[2]. B. Mekonen, and T. Bogale. "Experimental investigation of tensile, flexural and hardness properties of polyester resin echinatus fiber reinforced composite material." *ESM.* 2023, *11(2):* 151-162.

[3]. M. Y. Khalid, A. Al Rashid, Z. U. Arif, et al. “Recent advances in nano cellulose-based different biomaterials: types, properties, and emerging applications.” *J Mater Res Technol.* 2021, *14*, 2601-2623.

[4]. Eticha, Andinet Kumella, et al. "Biodegradable Biconstituent Melt-Blown Nonwovens for Air Filtration: Fabrication and Characterization." *Fibers and Polymers* 25.8 (2024): 2855-2873.

[5]. Tarekegn, Abrham Gebre. "Deformational Behavior of Fiber Reinforced Cement Based Materials Under Repeated Loading." *Zede Journal* 42.1 (2024): 93-104.

[6]. A. A. Betelie, A. N. Sinclair, M. et. al., “Mechanical properties of sisal-epoxy composites as functions of fiber-to-epoxy ratio.” *AIMS Mater Sci* 2019, *6(6):*985–996.

[7]. V. Thakur, A. Guleria, S. Kumar, et al., “Recent advances in nano cellulose processing, functionalization and applications: A review.” *Mater Adv* 2021, *2(6),* 1872-1895.

[8]. S. Mateo, S. Peinado, F. Morillas-Gutiérrez, et. al., “Nanocellulose from agricultural wastes: Products and applications—A review.” *Processes* 2021, *9(9),* 1594.

[9]. Angaw Chaklu, Araya Abera and Daniel Tilahun. "Extraction and characterization of Nanosilica particles to enhance mechanical properties of general-purpose unsaturated polyester resin" *SECM.* 2024, *vol. 31*, *no. 1*, pp. 20240001.

[10]. Pandurangan, Mohan Turup, and Krishnan Kanny. "Study of curing characteristics of cellulose nanofiber-filled epoxy nanocomposites." *Catalysts* 2020, *10 (8)*: 831.

[11]. T. Ramakrishnan, M. D. Mohan Gift, S. Chitradevi, et al., "Study of numerous resins used in polymer matrix composite materials." *Jhon Wiley & Sons* 2022, *2022(1),* 1088926.

[12]. A. Engidaw, A. Betelie, and D. Redda. "Experimental determination of strain in thin aluminum plate with central hole subjected to far-field tensile loading using digital image correlation (DIC)." *Engineering Solid Mechanics* 2024, *12(4)*: 423-436.

[13]. G. D. S. Maradini, M. P. Oliveira, G. M. D. S. Guanaes, et al., «Characterization of polyester nanocomposites reinforced with conifer fiber cellulose nanocrystals." *Polymers* 2020, *12(12),* 2838.

[14]. A. A. B. Omran, A. A. Mohammed, S. M. Sapuan, et al., "Micro-and nanocellulose in polymer composite materials: A rview." *Polymers* 2021, *13(2),* 231.

[15]. H. S. A. Embirsh, I. Stajčić, J. Gržetić, et al., "Synthesis, characterization and application of biobased unsaturated polyester resin reinforced with unmodified/modified biosilica nanoparticles." *Polymers* 2023, *15(18),* 3756.

[16]. H. Mohit, G. H. Kumar, M. R. Sanjay, et al., "Hybrid nanocomposites based on cellulose Nano crystals/nanofibrils: From preparation to applications." *Publ Ser Biomater* 2021, 223-245.

[17]. P. Thomas, T. Duolikun, N. P. Rumjit, et al., "Comprehensive review on Nano cellulose: Recent developments, challenges and future prospects." *J Mech Behav Biomed Mater 2020*, (*110)* :103884.

[18]. Cho, Mi-Jung, and Byung-Dae Park. "Tensile and thermal properties of Nano cellulose-reinforced poly (vinyl alcohol) Nano composites." *J Ind Eng Chem* 2011, *17(1)*: 36-40.

[19]. Abdullah, Nur Athirah, et al., "Nanocellulose from agricultural waste as an emerging nanotechnology material for nanotechnology applications–an overview." *Polimery* 66.3 (2021): 157-168.

[20]. Hailu, Hailemariam Nigus, and Daniel Tilahun Redda. "Prediction of Mechanical and Toughness Properties of Ni-Modified Cr-Mo Alloy Steels for Transmission Gear." *International Journal of Engineering Research in Africa* 56 (2021): 1-15.

[21]. Stepanova, Mariia, and Evgenia Korzhikova-Vlakh. "Modification of cellulose micro-and nanomaterials to improve properties of aliphatic polyesters/cellulose composites: A review." *Polymers* 2022, *14 (7)*: 1477.

[22]. M. Davallo, H. Pasdar, and M. Mohseni. "Mechanical properties of unsaturated polyester resin." *Int J Chemtech Res* 2010, *2(4)*: 2113-2117.

[23]. D. Trache, A. F. Tarchoun, M. Derradji, et al., "Nano cellulose: from fundamentals to advanced applications." *Front Chem 2020*, (*8*): 392.

[24]. Sudarman, Susilo, and Muhammad Taufik. "Synthesis and application of nano-silicon prepared from rice husk with the hydrothermal method and its use for anode lithium-ion batteries."

Mater. Sci. Energy Technol 2024, (*7*): 1-8.

[25]. Islam, Mohammad Tajul, Mohammad Mahbubul Alam, and Marina Zoccola. "Review on modification of Nano cellulose for application in composites." *Int. J. Innov. Res. Sci. Eng. Technol*. 2013, *2*(*10*) 5444-5451.

[26]. A. G. N. Sofiah, J. Pasupuleti M. Samykano, et al., « Harnessing Nature's Ingenuity: A Comprehensive Exploration of Nanocellulose from Production to Cutting-Edge Applications in Engineering and Sciences." Polymers, 2023, *15(14):* 3044.

[27]. R. R. Manikumar and T. Nageswara. "Comparative FTIR analysis of liquid and cured forms of ortho-unsaturated polyester resin and flame-retardant resin: Investigating chemical structure transformations." *Research Institute of Petroleum Processing* 2024, *24* (*1*): 91-116.

[28]. A. Zhadyra, B. Oral, I.Aziza and K. Aigul. "Modification of polymers to synthesize thermo-salt-resistant stabilizers of drilling fluids" *Open Engineering*, vol. 15, no. 1, 2025, pp. 20240097.

[29]. M. Kanda, S. Puggal, N. Dhall, & A. Sharma, "Recent developments in the fabrication, characterization, and properties enhancement of polymer Nano composites: a critical review." *Mater Today Proc* 2018, *5 (14)*: 28243-28252.

[30]. A. Mohammed, J. Hassan, S. Maan, and A-Q. Hussein. "Synergistic effect of nano-silica, steel slag, and waste glass on the microstructure, electrical resistivity, and strength of ultra-high-performance concrete" *Open Engineering*, vol. 15, no. 1, 2025, pp. 20240105.

[31]. A. A. Hibatallah, L. S. Rasheed, and W. A. Alaraji, "Experimental study on behavior of hybrid columns by using SIFCON under eccentric load" *Open Engineering*, vol. 15, no. 1, 2025, pp. 20240023.

[32]. M. Zenebe, E. Woldesenbet, and E. G. Koricho, "An experimental investigation on synthesizing and characterizing of modified bagasse reinforced natural rubber composite." *Journal of Composite Materials* 56.29 (2022): 4501-4514.

Performance Evaluation of Modified Solar Bubble Dryer for Red Pepper (Capsicum)

Mengistu Zelalem Zewdu[1,2*] and Kamil Dino Adem[1]

[1] *School of Mechanical and Industrial Engineering, College of Technology and Built Environment, Addis Ababa University, Addis Ababa, Ethiopia*
[2]*Department of Mechanical Engineering, Gambella University, Gambella, Ethiopia*
[*]*Corresponding author's E-mail address: habtemaryame21@gmail.com;*

DOI: https://doi.org/10.63990/zede.v43i.12976

ABSTRACT

Red pepper (Capsicum) is a spice type plant having major role in daily dish and plays an important role in the national economy of the country. One of the most serious challenges is getting the vegetable to market and selling before it spoils. The objective of this study was to design and manufacture modified solar bubble dryer to minimize and overcome the post-harvest loss of pepper vegetable crop. The experimental result showed that the efficiency of the dryer and the mass flow rate was found to be 22.6%, 3.53 $x10^{-4}$ *kg /s, respectively. It reduces the moisture content of the pepper from about 71% to 13% in 20 hours of operation. It lowers 20% drying time compared to an open-sun dryer. In addition to tis, the drying chamber temperature was higher compared to the ambient temperature by an average of 3 to 5°C.The logarithmic thin layer drying model fits for Red pepper with the respecting value of 0.96552, 0.00533 and 0.05003 values respectively.* It will take a minimum of 21 months to return the initial cost of the solar bubble dryer. *Therefore, it is recommended that the modular feature of the solar bubble dryer is an option to dry vegetables to reduce the alarming rate of postharvest losses in countries where drying vegetables is common.*

Keywords: Crop, Pepper, Performance evaluation, Solar bubble dryer, Vegetable,

1. INTRODUCTION

Solar drying, which uses the sun's energy to reduce moisture content in agricultural products, is an efficient and sustainable alternative to traditional drying methods. Solar drying techniques have proven effective for crops with high moisture content, such as red pepper which require careful drying to prevent spoilage and retain nutritional and sensory qualities. The traditional way of open-air drying is widely used due to its low cost, yet it exposes produce to environmental contaminants and leads to significant quality degradation due to fluctuating temperatures and humidity levels. In contrast, solar dryers offer a controlled environment that enhance drying speed, preserve nutrients, and reduce the risk of contamination [1].

Red pepper (Capsicum) is rich in bioactive compounds, including vitamins A and C, capsaicinoids, and carotenoids, which are susceptible to degradation under improper drying conditions. Research highlights the importance of temperature and humidity control to preserve these nutrients effectively. In addition to this, its color and flavor are used as the raw materials for different industries [2]. The antioxidant vitamins have a very important function to protect against oxidative damage and to prevent various diseases like cardiovascular disease and cancer [3].

One of the techniques to overcome post-harvesting loss is to implement a proper drying system. Drying is a means of reducing the weight, which serves to easily transport from place to place, provide longer shelf life, and have a small storage area [4]. However, much of the existing research has primarily explored direct sun

drying and conventional solar dryers. Studies specific to modified solar dryers designed for peppers are scarce, highlighting a gap in research on optimal drying methods that can preserve the unique nutritional profile of red pepper.

Drying red pepper in the sun requires a longer time and a large open area. The overall quality of the product may be affected by the availability of sunshine, insects, and fungal infestations. To get the best quality, the product needs to be dried in a shade. Pepper vegetable crop is 71% and 13% wet moisture content [5]. The traditional drying process of red pepper used to take around seven consecutive days, results in undesirable fermentation resulted in the market value of the product [6]. A solar dryer is one way of overcoming the problem of traditional open-sun drying.

The post-harvest loss of fruits and vegetables is estimated to be 30-40% in developing countries [7], and particularly about 20-30% of bell peppers and hot peppers are lost at the post-harvest stage every year [8]. These losses may be quantitatively measured by decreased weight or volume or can be qualitative, such as reduced nutrient value and unwanted changes to taste, color, and texture. The quantitative loss is caused by the reduction in weight due to factors such as spoilage and consumption by pests, as well as qualitative changes due to physical changes in temperature, moisture content, and chemical changes [9].

Inadequate methods of preserving agricultural produce, which may usually be produced in larger quantities during harvest, have remained one of the main problems of farmers. This loss is due to fungi attacks, insects, birds, rodent encroachment, and unpredictable weather effects such as dust, rain, and the wind, which have remained some of the inherent limitations of the traditional open-sun drying method widely being practiced by rural farmers [10]. The open-sun drying also results in physical and structural changes such as unnecessary shrinkage, case hardening, and loss of volatiles and nutrient components [11].

Different types of solar drying techniques have been tested for red pepper with distinct mechanisms for capturing and utilizing solar energy and degrees of success. Each method has distinct characteristics affecting product quality and drying efficiency.

Sun Drying: Traditional sun drying is widely used, especially in rural and low-resource settings. Although inexpensive, it exposes peppers to environmental contaminants, prolongs drying time, and often results in nutrient losses due to fluctuating temperature and humidity [12]. This method also produces inconsistent moisture levels, leading to higher risks of microbial growth and spoilage [13].

Conventional Solar Dryers: Solar cabinet dryers, among the most common solar dryers, enclose the product in a protected environment, reducing contamination and drying time. However, cabinet dryers often suffer from uneven heat distribution, which results in non-uniform drying and affecting product quality [14]. Noted that temperature control within cabinet dryers is often insufficient for consistent drying of high-moisture produce like peppers.

Hot-Air and Mechanical Drying: Mechanical dryers, which use electric or fuel-powered heat sources are faster and more effective in achieving uniform drying but are costly and energy-intensive. These dryers also risk nutrient degradation, as high temperatures can cause significant losses in bioactive compounds like vitamin C and carotenoids, particularly in sensitive produce like red pepper [15].

Solar Bubble Dryers: The solar bubble dryer is an innovative approach designed to address limitations in traditional solar drying systems by enhancing heat retention and drying uniformity. Using transparent plastic sheeting, the bubble

dryer traps solar energy creating a greenhouse effect that maintains stable internal temperatures.

While solar bubble dryers have shown promising results with staple crops, there is limited research on their application for moisture-sensitive, nutrient-rich crops like red pepper. Investigating the performance of modified solar bubble dryers for pepper drying could provide valuable insights into balancing drying efficiency with nutrient retention, therefore meeting both quality and economic demands.

Despite the advances in solar drying methods, there is a lack of comprehensive research specifically focusing on solar bubble dryers for drying red pepper. Most existing studies address general solar drying methods but do not adequately explore how modifications in solar bubble dryers could improve drying rates, nutrient retention, and uniformity for high-moisture crops like peppers. Moreover, there is a need to analyze the economic feasibility and energy efficiency of these modified dryers to assess their potential as a practical solution for small- to medium-scale farmers. The reported research showed that this type of dryer is economically acceptable and environmentally friendly [16].

This study aims to fill these gaps by evaluating the performance of a modified solar bubble dryer for red pepper drying using locally available materials in order to reduce thea post-harvest loss. The research focused on optimizing drying parameters to achieve rapid drying, high nutrient retention, and energy efficiency. By addressing these gaps, the study aimed to advance knowledge on solar drying technologies, contributing to the development of more effective and sustainable drying methods for sensitive, high-value crops of red pepper.

2. MATERIAL AND METHODS

The solar bubble dryer (SBD) used in the study had a drying floor, top cover, fabric mesh tray, DC fan, and solar panel that completely operated independently from fuel or the power grid. The following materials were used for the construction of the solar bubble dryer. The drying chamber was where the drying sample was placed and heated by the hot air that came from the collector. The drying chamber frame was constructed from RHS metal that had a thickness of 1 mm. The total area of the drying chamber was 2000 mm x 620 mm and had a height of 500 mm. The dryer had one tray used to dry the sample having a dimension of 2000mm x 56mm positioned at the center of the dryer. It was made of fabric mashed as shown in Figure 1 to avoid corrosion and contact between the drying sample and the metal frame as well as the plastic cover of the dryer. It had very tiny holes used to transfer solar radiation in order to have a uniform drying and heat-exchanging process between the upper and bottom surfaces of the dryer. This result increase improves both the performance and quality of the pepper. In addition, this solar panel may be used to drive the fan and helps to inflate the plastic top cover. The plastic cover enhanced the quality by preventing impurities like dust, insects, etc. as well as ensuring proportional air movement to achieve the objective of this research. The chimney was used to facilitate the moist air to leave the dryer by enhancing its pressure.

Figure 1 shows the 3D modeling of the solar bubble dryer. The main part that differs from the existing dryer was the fabric mesh tray, which is located at the center of the dryer. The fabric mesh tray enables uniform energy and air circulation in the drying product. In addition to this, the transparent top cover and the absorber make the dryer generate the energy on both sides. Chimney is the other modified main part of this dryer, which used to pressurize the exit air. It used to improve the moisture removal rate and facilitate the exit air to leave the dryer before it created dew. This helps to reduce the drying time

and to improve the quality and the uniform drying process of the product. As a result, it used to increase the efficiency of the dryer. The other modified part of this dryer is the bubble frame or the support of the top plastic cover, which helps to get the bubble shape of the dryer and carry the tray, which lets the drying product lay on it. In addition to this, it has a roller connection, which used to roll it in and out in the dryer. This makes the dryer portable and simpler to operate or handle by anybody.

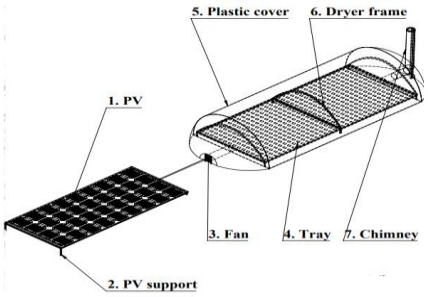


Figure 1 Design of solar bubble dryer developed using solid work.

Figure 2 shows the experimental setup of the dryer. The experimental set-up and the performance evaluation of the SBD dryer were conducted in the winter at AAiT mechanical engineering workshop. Addis Ababa, with a geographical location of latitude 9.2N and longitude 38.7E. The solar collector faced south with an east-west orientation and was tilted at 24.2°.

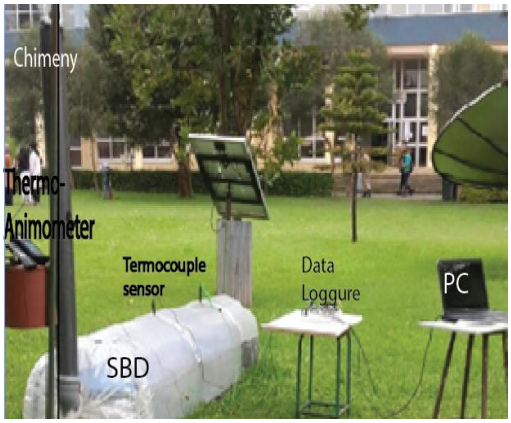


Figure 2 Experimental setup of SBD

Figure 3 shows the solar bubble dryer with respect to the convectional open sun dryers. As it is clearly seen that the pepper vegetable crop is exposed to open air and other impurities, the solar bubble dryer is protected from any type of external factors that cause damage or spoilage to the product. The materials used for the construction of the solar bubble dryer are listed in Table 1.

Figure 3 SBD and open sun drying

Table 1 Materials used for prototype model

No.	Part name	Material	Dimension
1	Collector and dryer cover	Plastic	4mx0.65mx0.5m
2	Dryer and tray support	(rectangular hallow steel) RHS	2mx0.62mx0.5m
3	DC Fan for air circulation	-	-
4	Solar panel	-	-
5	Fabric mesh or mosquito net	Tray	2mx.59m
6	Chimney	PVC	1.5m
7	Wheel for easily movement	Plastic made small wheel	-

2.1. Wet Basis Moisture and Dry Basis Moisture

Moisture content is one of the important parameters that determines the capacity of a solar bubble dryer. A drying product's moisture content is given either based on the total weight of the product to be dried or the amount of solid weight present in the product. It was calculated using the following equation [17].

$$MC(w.b)\% = \frac{M_w - M_d}{M_w} x100 \quad (1)$$

Where, M_w −weight of wet material

M_d- Weight of dry material

It is the other way of determining the amount of water present in the dried pepper vegetable crop. Even though it is not commonly applicable in most areas, it was found to be significant while calculating dry base-related calculations [18]. It can be easily determined using equation 2, once the percentage moisture content of the pepper is determined.

$$MC(d.b)\% = \frac{M_w - M_d}{M_d} x100 \quad (2)$$

2.2. Water to be Evaporated from the Drying Product

In the process of removing the moisture found in the pepper at the outer surface, it starts to evaporate, and the moisture content of the dried product found in the center dissipates and gets closer to the outer surface. This dehydration process depends on the porosity and surface area of the drying product. In addition to this, it will also be affected by the weather condition of the area where the pepper vegetable crop is being dried [18]. The amount of water content that could be removed was calculated using equation 3.

$$M_r = \frac{M_p(M_i - M_f)}{(100 - M_f)} \quad (3)$$

Where, M_P - the mass of the sample (kg)

M_i - primary vapor (% wb).

M_f - final vapor (% wb)

2.3. Moisture Removal Rate and the Amount of Air Required

The moisture removal rate is the amount of evaporated moisture over time that takes to completely remove the amount of moisture from the drying product [19].

$$DR = \frac{M_r}{t_d} \quad (4)$$

Where, M_r - the amount of moisture removed, (kg)

t_d - the time taken to dry the drying product, (hr)

The amount of air needed to evaporate the water content from the sample weight of the drying product can be determined using the equation explained by [20].

$$V_A = \frac{M_r(R_a - T_a)}{C_{pa}P_a(T_0 - T_f)} \quad (5)$$

Where, T_f= the final temperature at the end of the drying chamber

$$T_f = T_a + 0.25(T_c - T_a) \quad (5.1)$$

2.4. Heat Required to Remove Vapor Content

The amount of heat required to remove the moisture content from the drying product was evaluated using Eq. (6) [21]. It

depends on the amount of water removed from the drying product and the dehydration rate of the pepper vegetable crop [21].

$$Q_{load}=M_r L_v \quad (6)$$

Where, M_r- the dehydration rate of the pepper (kg)

L_v- the latent heat of water from the surfaces of pepper vegetable crop, 2320 kJ/kg

Q_{load}_amount of heat required to remove the moisture

2.5. The Amount of Mass Removed and Air Velocity

The amount of moisture removed from the drying product determines the duration to dry a particular amount of drying sample [22].

$$\dot{m}_a=\dot{\upsilon}_a\rho_a \quad (7)$$

Where $\dot{\upsilon}_a$-Volume flow rate of air

ρ_a-Density of air

$\dot{m}_a$- Amount of moisture removed from the drying product

The speed of exit air velocity is evaluated as the rate of discharge per the drying area of the dryer.

$$V=\frac{Q}{A} \quad (8)$$

Where, A- Area

Q- Discharge

V -speed of exit air velocity

2.6. Performance of collector, dryer and capacity of the dryer

The collector performance is expressed as the ratio of the amount of energy migrated from the collector to total solar incidence over the specified time [23].

$$\eta_c = \frac{v\rho\Delta TC_p}{I_c A_c \tau} x100 \quad (9)$$

It is the ratio of the amount of heat required to dry the drying product per the product of solar radiation, collector area, and time taken to dry the product. It is the main criteria to evaluate the performance of the given dryer [24].

Drying Efficiency,

$$\eta_d=\frac{M_w L}{I_c A_c t\tau} x100 \quad (10)$$

The capacity of the dryer can be determined once the volume of the drying product is determined. The volume of pepper vegetable crop may be obtained from equations given in [25] as:

$$\nu_y=W_T L_T Y_l \quad (11)$$

The total volume of the drying product on the tray may evaluated as;

$$\nu_T=n\nu_y=nW_T L_T Y_l \quad (12)$$

Where n- Number of tray

The relation between the bulk density and the total volume may be evaluated as follows:

$$W_w=\rho_b \nu_T \quad (13)$$

Where Y_l - Thickness of the pepper vegetable crop (m)

W_T- tray width (m)

L_T-tray length, (m)

M_p - the mass of the drying product (kg)

ρ_b- the bulk density of pepper vegetable, $402.1(kg/m^3$ [25]

$$L_d \cong L_T \quad (14)$$

2.7. Sizing of the Power of Fan and Determine the Capacity of the Solar Panel

The international standard of the conducive thermal comfort range for blowers found between 50 and 100cm. It can be related to that as $100cm=0.0491747m^3/s$ [26].

Fan horse power =

(Air flow×pressure)/(6320× efficency of fan) [24].

The efficiency of the fan ranged in between 70 and 85%. For this study, an axial fan that has the maximum efficiency was taken for adequate air circulation and optimum heat transfer rate in the drying chamber.

The capacity of a solar panel is based on the size of the panel. Panel generation factor and the size of the module determine the maximum watt that can be produced. In addition to this, the weather condition is also one of the determinant factors for the given solar panel. This

study considered a 20-watt axial DC fan that runs for 8 hours per day to supply sufficient air for the drying product. Based on this, the daily energy consumption can be evaluated as [27].

$$\text{Daily energy consumption} = \text{power x time} \quad (16)$$

For PV Module Sizing a factor of 2.5 is reasonable [28].

$$\text{Total PV module needed} = \frac{\text{Total dayily energy consumption}}{\text{PV corection factor}} \quad (17)$$

$$\text{No. of PV module} = \frac{\text{Total PV Module needed}}{\text{Standard PV Module}} \quad (18)$$

2.8. Design of chimney

The chimney helped to pressurize the exit air to live the drying chamber before it created dew. This ensures the proportional drying rate of the drying product throughout the dryer.

$$M_e = M_r + M_a \quad (19)$$

Where M_e- mass of exit air (kg)

The density of air that enters the dryer may be evaluated as;

$$\rho_i = \rho_o \times \frac{T_o}{T_a} \quad (20)$$

Were, ρ_i – Air density (kg/m^3)

ρ_o - Density of air at 0°C (kg/m^3)

T_a- Ambient air temperature (°C)

T_0- Exit temperature (°C)

The height of the chimney was assumed to be 2 m, and the upward draft was calculated based on the equation as follows:

$$D_P = (\rho_i - \rho_e) \times H \times g \quad (21)$$

Were, D_P-Produced draft (kg/ms^2)

ρ_i - Density of inlet air (kg/m^3)

ρ_e - Density of exit air (kg/m^3)

g- Gravity (m/s^2)

The actual draft is the ratio of the product draft per the exit velocity of the air. It may be evaluated as follows:

$$V_e D_a = 0.75 \times D_P \quad (22)$$

D_a- Actual draft (kg/ms^2)

The exit air velocity has to be evaluated as the ratio of the square root of two times the actual draft per the exit density, as illustrated in Eq. (24).

$$V = \sqrt{\frac{2D_a}{\rho_e}} \quad (23)$$

Where V- exit air velocity (m/s)

The rate of exit air may be evaluated as the ratio of the rate of air required per number of chimneys found in the given dryer;

$$Q_c = \frac{Q_a}{n} \quad (24)$$

Where, Q_a – rate of expected air needed (m^3/s)

n- Number of chimney (assumed to be 1)

Since exit air is discharged using n chimneys, once the number of chimneys is specified, the rate of exit air can be evaluated using the Eq. (25).

$$\dot{\upsilon}_{ch} = \frac{Q_a}{n} \quad (25)$$

Where $\dot{\upsilon}_{ch}$ - volume flow rate (m^3/s)

Therefore, the area of the chimney may be evaluated based on the amount of air interring in to the dryer and its speed in the dryer.

$$a_{ch} = \frac{Q_a}{V} \quad (26)$$

Where a_{ch} - Diameter of chimney (m)

The diameter of the chimney was evaluated based on the area of the chimney and determined as follows:

$$d_{ch} = \sqrt{\frac{4 \times a_{ch}}{\pi}} \quad (27)$$

Where d_{ch} - Diameter of chimney (m)

2.9. Sensor positioning of the dryer

The primary goal of this research was to address the problem of traditional ways of drying and improve the efficiency of the existing solar bubble dryer. Sensor position is one of the technical parameters to determine and analyze the distribution of the temperature, which originated from the collector of the dryer. Thermocouples were located in three different place a (inlet), b (middle), and c (outlet) as shown in Figure 4. Each thermocouple was connected with the data logger. Moreover, the data logger receives and transfers the data to the PC for further analysis.

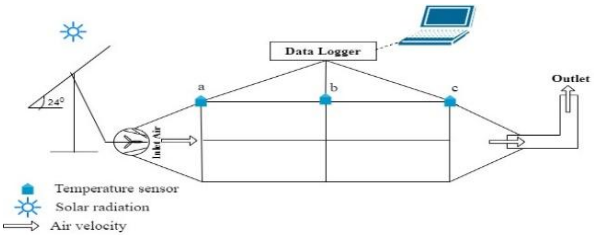


Figure 4 Sensor positioning and layout of SBD experimental set up

Figure 4 shows the positioning of the sensors Red dots represent the instrument used to measure the temperature integrated with the instrument called National Instrument NI cDAQ-9172.

3. RESULTS AND DISCUSSION

3.1. Moisture Removal Rate of SBD

Figure 5 shows the moisture removed from the sample which was measured in two ways: by using a device called a moisture meter, and by using an analytical balancing machine or analytical weighting scale. Moisture content of pepper was measured in both open-sun and solar bubble dryer methods in order to understand the difference between them by recording the result on the data spread sheet every 30-minute interval. The moisture content of the pepper did not build up every next day since it was packed in zipped plastic bag, which did not let it lose or gain moisture overnight. Moreover, zipped plastic was stored in a desiccator. As it is clearly seen in Fig.5, the final moisture contents of the sample pepper dried in the solar bubble dryer and the traditional type dryer were 13.6% and 17.7%, respectively, with a weight of 7,2233 g and 10.6163, respectively.

Figure 5 Final moisture content of pepper vegetable crop on both SBD and OSD

3.2. Drying performance

Figure 6 shows that drying performance were observed to be lower in the winter season than in the summer season due to the effects of rain, cold, and foggy weather, as well as low temperatures. [29]explained that the drying rate of the dryer gets higher when the drying air temperature provided by the solar collector gets high. This is due to the intensive heat and mass transfer followed by a high rate of water evaporation.

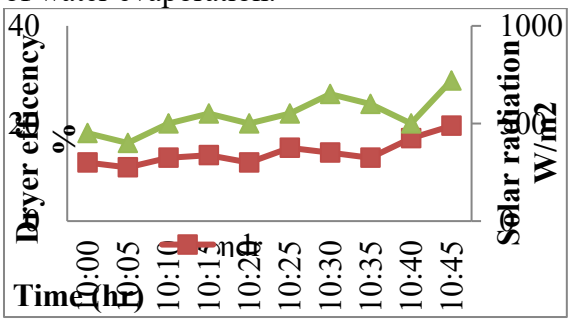


Figure 6 Dryer efficiency vs. solar radiation variation

Drying of red pepper in solar drying reduced the moisture content from approximately 71% to approximately 13% in 20 h. At the average solar radiation of 356.3 W/m2 and mass flow rate of 3.53 x10^{-4} kg /s, and the collector and drying system efficiencies were 15% and 22.6%, respectively. Even though the initial moisture content of the pepper vegetable crop and rice are too different, this dryer efficiency is much more efficient and faster than what it was reported in [30].

3.3. Moisture Distribution

The moisture distribution of the pepper vegetable crop was faster than that of the open-air dryer. The weight loss was recorded every half-hour interval in order to realize the reduction weight of the drying sample. As it is known, if the drying process takes more than one day, the drying product may gain moisture content in the dormant time [31]. However, this problem was resolved using the equipment called a dissector that helps to contain the moisture or weight of the sample until the next experiment. This equipment resolves the re-watering phenomena of the pepper vegetable crop and helps accurately conduct the experimental process of the drying product.

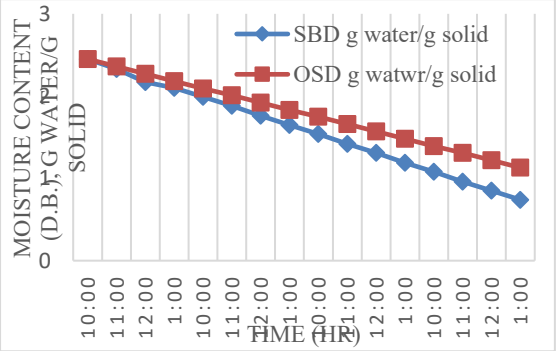


Figure 7 Moisture content (dry basis) with time of pepper vegetable crop for SBD & OSD

3.4. Drying Kinetics Models

The most used type of drying kinetics model was implemented in this study to determine the best one for drying pepper vegetable crops under a solar bubble dryer. This study was based on the maximum result and minimum value to determine the best type of drying kinetics for the drying product. Among them, the logarithmic model satisfied the requirement to be selected as the tine layer model for the drying process, and related results were reported in [32].

Table 2 Drying kinetics models of pepper vegetable crop

Model no	Model constant	x^2	R^2	RMSE
1	k=0.05837	0.00837	0.88904	0.09149
2	a=0.94208, k=0.05531	0.00824	0.89229	0.09076
3	**a=0.98796, k=0.04359, c=0.08125**	**0.00533**	**0.96552**	**0.05003**
4	a=0.47104, k_0=0.0552972, b=0.47104, k_1=0.0553017	0.00847	0.89229	0.09203
5	k=0.00688, n=1.164944	0.00817	0.90123	0.07028
6	a=0.01023, b=0.00025	0.00824	0.90032	0.08973

3.5. Payback Period

Payback period is one of the simplest methods to find out the period by which the investment on the project recovered from the net cash inflows, i.e., gross cash inflow less the cash outflows. In short, it was defined as the period required to recover the original investment cost. The payback period starts with a preconceived notion that the management wants to recover the cost of investment within a specific period. It is found that it may take a minimum of 21 months to return the initial cost of the solar bubble dryer. It is seen that the solar bubble dryer is feasible technically and economical when it is compared to a related dryer, which was reported in [33].

4. CONCLUSION

Proper drying of pepper vegetable crops has been one of the major problems in Ethiopia. The performance evaluation of the SBD dryer was conducted in the CoTBE mechanical engineering workshop in the winter season. The evaluated output indicated that the pepper vegetable crop moisture reduced from about 71 to 13% in 20 hours of operation. The temperature distribution and moisture removal rate were uniform in all locations of the SBD

dryer during drying time. The total drying time of the dryer was 20 hours, saving 20% of the drying time of the pepper vegetable crop when it was compared with open sundries. The drying rate as well as the dryer and the collector efficiency were found to be 3.53 $x10^{-4}$ kg /s, 22.6%, 15%, respectively. The logarithmic thin layer drying model fits better for pepper vegetable crops than other models based on the value of R^2, x^2andRMSE. The analysis showed 0.96552, 0.00533, and 0.05003 values, respectively. It will take a minimum of 21 months to return the initial cost of the solar bubble dryer. It was economical and feasible when compared to similar dryers. Moreover, the technical performance and the experimental observation resulted in the finding that the pepper vegetable crop sample used to dry in the dryer has better quality, quantity, and proportional drying profile than open-sun drying. Therefore, it was concluded that the modified solar bubble dryer can be one of the options for drying pepper vegetable crops. It is recommended that this dryer is an option to dry vegetables to reduce the alarming postharvest losses in the country.

CONFLICT OF INTEREST

The authors declare that they don't have any conflict of interest.

ACKNOWLEDGEMENTS

The authors would like to thank the School of Mechanical & Industrial Engineering, Addis Ababa Institute of Technology, Addis Ababa University, and Gambella University for funding the present study and the technical staff for constructing the modified solar bubble dryer.

REFERENCE

[1] S. Admassu, (2015). "Post-harvest sector challenges and opportunities in Ethiopia" *Food Technologist, Ethiopian Agricultural Research Organization Addis Ababa, Ethiopia*.

[2] K. Esayas, A. Shimelis, F.,Ashebir, et al. (2011). "Proxmate Composition, Mineral Content and Antinutritional Factors of Some Capsicum (Capsicum annum) Varieties Grown in Ethiopia" *Chemical Society of Ethiopia, 25*(3), 451-454.

[3] W. Gobie, (2019). "A seminar review on red pepper (Capsicum) production and marketing in Ethiopia." *Cogent Food & Agriculture, 5*(1). doi:1647593

[4] S. B. Mansour-Gueddes, N. Tarchoun, J. A. T. Silva, and S. Saguem, (2010). "Agronomic and Chemical Evaluation of Seven Hot Pepper (Capsicum annuum L.) Populations Grown in an Open Field." *Fruit, Vegetable and Cereal Science and Biotechnology Global Science Books, 4*(1), 93-97.

[5] A. A. E. Sebaii, and S. M. Shalaby, (2012). "Solar drying of agricultural products." *Renewable and Sustainable Energy Reviews, 16*, 37–43.

[6] A. G. M. B. Mustayen, S. Mekhilef, and R. Saidur, (2014). "Performance study of different solar dryers." *Renewable and Sustainable Energy Reviews, 34*, 463–470.

[7] R. Mussema, (2006). "Analysis of red pepper marketing: the case of Alaba and Siltie in SNNPRS of Ethiopia*."* M. Sc. Thesis Agricultural Economics, Haramaya University.

[8] J. Singh, (2015). "Review Paper of Study on Solar Dryer." *Journal of Mechanical and Mechanics Engineering, 1*(1), 14-24.

[9] H., Asl. Darvishi, A. R., Asghari, A., Azadbakht, M., et al. (2014). "Study of the drying kinetics of pepper." *Journal of the Saudi Society of Agricultural Sciences, 13*, 130-138.

[10] A. Fudholi, M. Y. Othman, M. H., Ruslan, and K. Sopian, (2013b). "Drying of Malaysian Capsicum annuum L. (Red Chili) Dried by Open and Solar Drying." *Hindawi Publishing Corporation International Journal of Photoenergy*, 9.

[11] A. Fudholi, K. Sopian, M. H. Ruslan, M. A., Alghoul, and M. Y. Sulaiman, (2010). "Review of solar dryers for agricultural and marine products." *Renewable and Sustainable Energy Reviews, 14*(1), 1-30. doi:10.1016/j.rser.2009.07.032

[12] D. G. Mercer, (2014). "An Introduction to the Dehydration and Drying of Fruits and Vegetables": Departement of Food Science University of Guelph Ontario, Canada.

[13] B. O. Bolaji, (2005). "Performance Evaluation of a Simple Solar Dryer for Food Preservation." *The 6th Annual Engineering Conference Proceedings*.

[14] M. A., Al-Neama, and I. Farkas, (2016). "Influencing of solar drying performance by chimney effect." *Hungarian Agricultural Engineering, 30*(11).

[15] A. K., Kamble, I. L. Pardeshi, P. L., Singh, et al. (2013). "Drying of chilli using solar cabinet dryer coupled with gravel bed heat storage system." *Journal of food research and technology, 1*(2), 87-94.

[16] Mussema, R. (2006). "*Analysis of Red Pepper Marketing: the Case of Alaba and Siltie in SNNPRS of Ethiopia."* M. Sc. Thesis Agricultural Economics, Haramaya University.

[17] , F. K. Forson, M. A. A. Nazha, F. O. Akuffo, and H. Rajakaruna, (2007). "Design of mixed-mode natural convection solar crop dryers: Application of principles and rules of thumb." *Renewable Energy, 32*, 2306-2319.

[18] K. S. Tonui, E. B. K. Mutai, D. A. Mutuli, D. O. Mbuge, and K. V. Too, (2014). "Design and Evaluation of Solar Grain Dryer with a Back-up Heater." *Research Journal of Applied Sciences, Engineering and Technology, 7*(15), 3036-3043.

[19] B. O. Bolaji, (2005). "Performance Evaluation of a Simple Solar Dryer for Food Preservation. "*The 6th Annual Engineering Conference Proceedings*.

[20] M. Leon, (2002). "Renewable and Sustainable Energy Reviews." *Renewable and Sustainable Energy Reviews, 6*, 367-393.

[21] S. Abubakar, (2017). "Design, simulation, construction and performance comparison of mixed mode solar crop dryers with and without thermal sto rage".

[22] Ü. Necmiye, A. Kılıçkan, and I. Yalçın, (2010). "Effects of moisture content on some physical properties of red pepper (Capsicum annuum L.) seed." *African Journal of Biotechnology, 9*(24), 3555-3562.

[23] S. O. Ismail, S. J. Ojolo, J. I. Orisaleyeand F. A. Olusegun, (2014). "Design of an Office Table Solar-Dc Powered Fan." *Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS), 5*(1), 1-5.

[24] M. B. Ibrahim, A. Mohammed, and H. Abubakar (2016). "Development of a Solar Powered Standing Dc Fan Using Three Phase." *American Journal of Engineering Research (AJER), 5*(12), 148-154.

[25] B. Axtell, *Small Solar Tunnel Dryer*.

[26] S. Romuli, S. Schock, M. K. Somda, and J. Müller, (2020). "Drying Performance and Aflatoxin Content of Paddy Rice Applying an Inflatable Solar Dryer in Burkina Faso." *Applied sciences, 10*(3533). doi:10.3390/app10103533

[27] M. S. Seveda, (2012). "Design andDevelopment of Walk-In Type Hemicylindrical Solar TunnelDryer for Industrial Use." *International Scholarly Research Network ISRN Renewable Energy*, 9.

[28] M. S. Hassan, M. S. Islam, M. A., Hossain, and M. S. Rahman, (2012). "Development and performance evaluation of hybrid dryer for quality grain seeds." *International Journal of Energy Machinery, 5*(1), 42-52.

[29] M. A. Alam, C. K. Saha, M. M. Alam, et al., (2020). "Experimental investigation of solar bubble dryer for rough rice drying in Bangladesh." *Journal of Bioscience and Agriculture Research, 23*(02), 1920-1930.

[30] A. Fadhel, S. Kooli, A. Farhat, and A. Belghith, (2014). "Experimental Study of the Drying of Hot Red Pepper in the Open Air, under Greenhouse and in a Solar Drier." *International Journal of Renewable Energy & Biofuels, 2014*, 1-14.

[31] M. S. Seveda, (2012). "Design andDevelopment ofWalk-In TypeHemicylindrical Solar TunnelDryer for Industrial Use." *International Scholarly Research Network ISRN Renewable Energy*, 9.

[32] R. Patil, and R. Gawande, (2016). "A review on solar tunnel greenhouse drying system." Renewable and Sustainable Energy Reviews, 56, 196–214.

[33] M. Mid, A. R. Yadollahinia, and S. Rafiee, (2006). "A thin-layer drying model for paddy dryer." Proceedings of the international conference on Innovations in Food and Bioprocess Technologies.

Effect of Thermal Annealing on Device Performance of Organic and Solid-state Dye-sensitized Photovoltaics

Desta Gebeyehu[1,*]

[1]*Department of Physics/DTE CELS, Addis Ababa University, Addis Ababa, Ethiopia*
**Corresponding Author's E-mail address: dgebeyeh68@gmail.com*

DOI: https://doi.org/10.63990/zede.v43i.12977

ABSTRACT

In this study, the impacts of thermal treatment on the photovoltaic properties of emerging 3rd generation photovoltaics based on polymer organic and solid-state dye-sensitized solar cells are investigated. We have demonstrated a polymeric organic and solid-state dye-sensitized emerging photovoltaic device with power conversion efficiencies of over 3% under standard AM 1.5 illuminations at 80 mW/cm^2 as well as 100 mW/cm^2 using optimized thermal annealing condition at 80 ^{0}C for 10 minutes. Analysis of experimental results indicates that thermal annealing conditions could reduce the density of defect states, promotes carrier transport, increases photovoltaic characteristic values, and ultimately improve the performance of these photovoltaics. To enhance the maximum photovoltaic performance of these solar cells, soaking in the solvent, followed by spin-coating the P3HT solution and thermal annealing treatment at optimized thermal annealing processing temperature, proved to be a useful technique with power conversion efficiency of 3.6%. Consequently, we found that soaking in the solvent, spin-coating the P3HT solution, and thermal annealing the solid-state dye-sensitized photovoltaic device at 536 nm with a power intensity of 0.1 mW/cm^2 produced an incident photon to collected electron efficiency (IPCE) of 72%.

Keywords: Efficiency. Organic and solid-state dye-sensitized Photovoltaics, Optimization; Thermal annealing mechanisms;

1. INTRODUCTION

Within the next 50 years, the world's energy demand is anticipated to double. However, as fossil fuels become harder to come by, they are blamed for the rise in atmospheric carbon dioxide, which is the primary cause of greenhouse gas emissions and environmental harm. Hence, developing environmentally friendly, renewable energy is one of the challenges to society in the 21st century to mitigate climate change, to reduce greenhouse gas emissions, and increase the security of our global energy supply system. One of the renewable energy technologies is photovoltaics (PV), the technology that directly converts sunlight into electricity through a process called the photovoltaic effect. PV is one of the most well-known and the fastest growing of all the renewable energy technologies at present. The development of affordable solar photovoltaics is one of the most promising long-term solutions to keeping the CO_2 concentration in the atmosphere at safe levels. To date, one of the most significant sources of energy that does not hurt the environment and has the huge potential to replace fossil fuels as a source of energy is solar energy. However, one barrier impeding the expansion of large-scale power source application of solar photovoltaic systems has been the high price of the solar cell module due to high production costs of solar-grade silicon. In order for photovoltaic technology to be made truly competitive with fossil fuels, large-scale manufacturing of these devices will need to be economically and technically viable in order to supply a significant amount of the global energy

needs. This will offer a pathway to a sustainable energy source.

Nowadays, the most popular kind of solar cell is a crystalline silicon cell, which can be composed of silicon that is either monocrystalline or polycrystalline. Although they can be expensive to produce, they are effective and long-lasting. Through ongoing research and development, these cells' efficiency has grown dramatically over time and can now reach up to a high power conversion efficiency (PCE) of 26.7% [1], whereas multicrystalline silicon solar cells that are now on the market have power conversion efficiency of between 14 and 19%. On the other hand, recently developed bulk heterojunction organic as well as dye-sensitized photovoltaics “emerging third generation photovoltaics“ could exceed the limits of single junction devices and lead to ultra-high efficiency for the same production costs of first and second generation photovoltaic devices [2]. Several promising technologies for those emerging third-generation low manufacturing cost photovoltaics are currently available and grouped under polymer based organic solar cells [3, 4, 5, 6, 7], small molecule-based organic solar cells [8, 9, 10], dye-sensitized solar cells [11, 12, 13, 14, 15], and solid-state dye-sensitized solar cells [16, 17, 18, 19, 20].

In recent years, researchers have developed organic photovoltaic cells that are more effective, stable, and economical. For solar cell applications, power conversion efficiency (PCE) is a crucial characteristic that can be increased by device engineering, material innovation and tailoring the active layer materials. Emerging third-generation photovoltaic technologies have recently attracted significant interest from industry and academic researchers and demonstrate distinct advantages over their inorganic counterparts due to their lighter weight, potential use in flexible devices, small thickness, inexpensive materials, ability to go through a solution process, and low cost in large-scale industrial production such as doctor-blading [21], spray coating and roll-to-roll processing [22, 2]. While roll-to-roll manufacturing is a potential solution for large-scale production, its practical development and integration into the organic solar cells manufacturing process require further research and discussion. Additionally, emerging third generation photovoltaic technologies are able to produce more power from the same quantity of sunshine than first generation (i.e. silicon-based) solar cells and second generation (i.e. thin films) solar cells due to their higher efficiency levels. They so represent a desirable alternative for both large- and small-scale solar energy applications.

The key property which makes organic photovoltaics so attractive is the potential of roll-to-roll processing on low cost substrate with standard coating and printing processes. However, several challenging scientific and technological problems remain to be solved. Both UV-induced photon exposure and environmental exposures from oxygen and moisture cause organic solar cells to degrade quickly, changing the efficiency of power conversion. It has been shown that the fundamental material properties, such as mobility of holes, defect densities in the absorbing material, and defect densities at the interface change as the device is exposed to air and to environmental exposure. So, the main challenges that today still hamper widespread commercialization are in the areas of cell and module lifetime, reliability, encapsulation barrier and volume manufacturing techniques. It should be noted that bigger area devices currently have a power conversion efficiency of 3.5% to 12% [2, 23], which is significantly lower than other types of devices. Additionally, roll-to-roll manufacturing is being developed and tested for organic photovoltaic devices on a lab scale,

including printing or coating methods such screen, inkjet, offset, gravure, slot die, spray, and others [24, 25, 26, 2]. However, for p-i-n type small molecule based organic vacuum deposited tandem solar cells [27], extrapolated lifetimes corresponding to more than 30 years of sun illumination have been achieved. On the other hand, after 72 h of thermal annealing at 150 °C, stable power conversion efficiency of polymeric solar cell as high as 4.7% was maintained [28].

Various architectures for organic photovoltaic devices have recently been researched. Generally speaking, four crucial processes must be optimized for an organic-based photovoltaic cell [29, 2] to be successful in converting solar energy into electrical energy: (1) light absorption followed by excitons generation; (2) excitons diffusion towards active interface (i.e. charge transfer) and creation of separated charges at the donor - acceptor interface; (3) selective transport of the charges through the bulk of the device; and (4) charge collections by the appropriate dissimilar metallic electrodes (see Figure 1). Figure 1 depicts bulk heterojunction organic photovoltaics that are competitive. The photogenerated charges must migrate to these electrodes without recombining after the charge transfer reaction. Finally, it is important that the photogenerated charges can enter the external circuit at the electrodes without interface problems. The photoactive organic layer's absorption spectrum must match the solar emission spectrum in order for the layer to effectively gather photons, and it must be thick enough to completely absorb input light. It is also possible to improve light harvesting by using narrow band gap polymers; however, it can be difficult to find a polymer that can have a significant amount of overlap with the solar emission spectrum. Additionally, previous studies have demonstrated that optimizing the active layer materials by taking into account band gap, mobility, solubility, morphology, molecular energy level, photon absorption, and other factors is the key to increasing the solar cell efficiency [30].

Scheme of the operation of bulk heterojunction organic photovoltaics is shown in Figure 1. In this kind of device, a blend of a donor and an acceptor organic material is present in the active layer. The lowest unoccupied molecular orbital (LUMO), also known as the conduction band, and the highest occupied molecular orbital (HOMO), also known as the valence band, in one material and the corresponding HOMO and LUMO in the other material, differ energetically in an organic semiconductor. The chemical structure, self-ordering, concentration and mixing ratio of the polymer and fullerene, the impacts of the fullerene on the polymer, thickness, concentration of the solution, and heat and vapor annealing are some of the factors that may affect the nano-morphology of the active layer. The effective operation of solar devices is demonstrated experimentally to need optimal morphology [31]. To increase the performance of the solar cell device, the active layer materials must be tailored taking into account factors like band gap, mobility, solubility, morphology, molecular energy level, photon absorption, and so on.

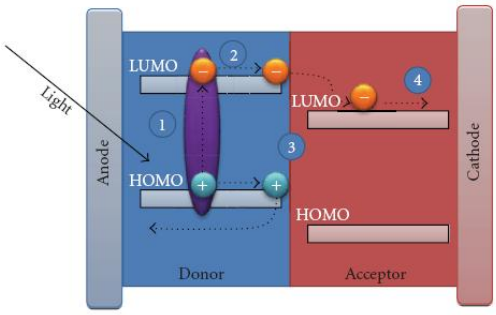


Figure 1. Scheme of operation principle of bulk heterojunction organic photovoltaics

The optimization of bulk heterojunction polymer organic and solid-state dye-sensitized solar cells has a number of unmet research needs. Therefore, it is still difficult to achieve high power conversion efficiencies (PCEs) in solid-state dye-

sensitized solar cells as well as polymer-fullerene solar cells. In order to improve PCE, more study is required to optimize device architecture, material properties, manufacturing methods, or device processing conditions. Therefore, in this work in particular, the thermal annealing of emerging 3rd generation photovoltaics based on bulk heterojunction polymer organic and solid-state dye-sensitized solar cells were examined using material properties, manufacturing methods, or device processing conditions. Titanium dioxide (TiO_2) nanoparticles arranged in a porous layer, a dye monolayer covering the particles, and a "hole-transporter" that enters the pores of the TiO_2 layer is the three main components of dye-sensitized solar cells (DSCs). The main research challenges organic-based as well as hybrid solid-state dye-sensitized photovoltaic devices relate to the much needed increase in high power conversion efficiency, stability improvement (through better encapsulation and material degradation prevention) and the development of an adapted low-cost, large area deposition techniques [24, 25, 32] for roll-to-roll manufacturing technologies [26, 33], such as ink-jet printing, doctor-blade coating, screen printing, and spray coating.

The overall energy conversion efficiency, η_e of the photovoltaics cell has been performed using the equation

$$\eta_e = (V_{oc} x I_{sc} x FF) / P_{inc} \quad (1)$$

under different intensities, where V_{oc}, I_{sc}, FF and P_{inc} are the open circuit potential, short circuit current, fill factor and incident light power, respectively. I_{sc} is the current that is recorded when the externally applied voltage is zero. I_{sc} represents the number of charge carriers that are generated and eventually collected at the electrode at a short circuit condition. Enhanced optical/electrical parameters such as small band gap, high absorption coefficient, smaller phase separation, and high carrier mobility improves I_{sc}. V_{oc} is the open circuit potential when the current density output is zero. V_{oc} has been reported to be mainly dependent on the work function difference of metal contacts. If an ohmic contact is formed at the electrodes, V_{oc} is dependent on the HOMO-LUMO difference between the donor and acceptor molecules. The fill factor defines the shape of the I-V curve and we determine the value of the fill factor of the device, FF, by calculating the area of the maximum power rectangular area under the I-V curve in the 4th quadrant. Therefore, the fill factor is defined as

$$FF = (V_{\max} x I_{\max}) / (V_{oc} x I_{sc}) \quad (2)$$

where V_{max} and I_{max} are voltage and current at the point of maximum output power of the rectangle with the intersection of the current-voltage curve respectively. FF represents the dependence of current output on the internal field of the device and is quantified by series resistance and shunt resistance. A typical current-voltage curve of photovoltaic device is shown in Figure 2.

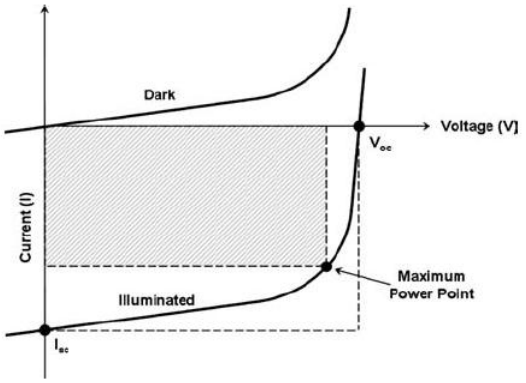
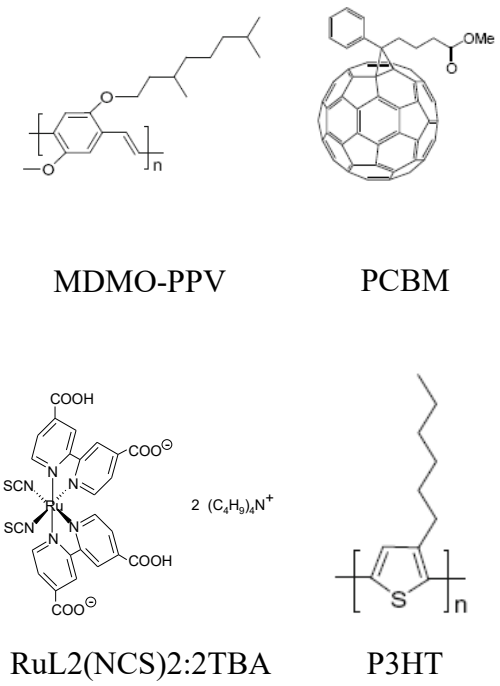


Figure 2. Graph of current versus voltage for photovoltaic devices. The figure shows how the device characteristics change upon illumination [34].

2. MATERIALS AND METHODS

Over the past few decades, research on organic semiconducting materials has become a subject rich in fundamental knowledge of unusual electrical phenomena and photo physics. The creation of organic photovoltaic

technology is one application of this fundamental research [2]. In order to improve the efficiency of power conversion, researchers put a lot of work and research into developing potential narrow band gap polymers, dyes, and semiconducting materials [30]. The absorption of the organic photovoltaic (OPV) materials should match the indoor light spectrum and should also have high external quantum efficiency (EQE) in order to produce high-efficiency OPV under the indoor lighting conditions. The chemical compositions of the organic materials employed in this investigation are shown in Figure 3.



MDMO-PPV PCBM

RuL2(NCS)2:2TBA P3HT

Figure 3. The chemical structures of the materials used in this study [5, 20].

All of these organic materials used in this study, such as TiO_2 paste and the solubilized C_{60} derivative [6,6]-Phenyl C61-Butyric Acid Methyl Ester (PCBM), poly (3-hexylthiophene), poly [2-methoxy, 5-(3',7'-dimethyl-octyloxy)]-p-phenylene vinylene (MDMO-PPV), were purchased from commercial sources. The use of new materials has enabled significant increases in bulk heterojunction organic solar cells as well as dye-sensitized solar cell efficiencies. Materials and methods [11, 12, 7, 16, 14] are a critical factor for both efficiency and lifetime.

The semi-transparent ITO coated glass substrates with a sheet resistance of ~15Ω/□ used in the fabrication of the polymeric bulk heterojunction organic solar cell devices described here were cleaned in ultrasonic baths of acetone, methanol, and isopropanol before being treated with oxygen plasma. ITO was spin-coated to a thickness of 100 nm with Poly(ethylene dioxythiophene) doped with Polystyrene Sulphonic Acid (PEDOT:PSS, Bayer AG), which enhances and stabilizes the electrical contact between the transparent anode (ITO) and the polymer active layer. Additionally, it reduces the substrate's surface roughness. A mixed composite of donor and acceptor bulk heterojunction organic photovoltaics devices was produced after the PEDOT: PSS film had dried overnight. Thus, the PEDOT: PSS was spin-coated with a photoactive layer made of MDMO-PPV: PCBM (1:4 by weight percent ratio) that had a thickness of 100–120 nm from chlorobenzene solution. MDMO-PPV can be used as a good donor material and PCBM is suitable for accepting charge carriers. LiF and Al were deposited in two layers for the cathode. First, a thin layer of LiF with an average thickness of 0.6 nm was thermally deposited onto the active layer (10^{-6} Torr) from a tungsten boat. Finally, a shadow mask was used to define a device region of 3 mm^2 to 7.5 mm^2 before a 100 nm aluminum cathode layer was thermally placed on top of the LiF. In an argon-filled dry box, device fabrication was carried out. The finished devices were thermally annealed by being placed directly on a hotplate with a digital temperature control in a glove box filled with nitrogen gas. The devices were annealed, then placed in a glove box and cooled to room temperature before the measurements were performed.

On the other hand, hybrid solid-state DSCs were made using ITO or SnO_2:F glass

substrates with a sheet resistance of ~15Ω/□ with a 5 mm^2 active area and were constructed using the methods that have been described in the literature [16]. A hybrid solar cell is made up of both organic and inorganic components in this instance. Moreover, inorganic semiconductors can be made as nanoparticles and inorganic semiconductor nanoparticles have the advantage of having high absorption coefficients and size tenability (i.e. functions as an excellent photoanode). However, polymeric hole conducting materials are a viable alternative to liquid electrolyte in DSCs due to their lower cost and ability to be chemically tailored to meet a range of technological requirements. Contrary to the liquid cell, the hole transporting material could pass through the pores of the TiO_2 to make an ohmic contact with the glass substrate coated with transparent conducting oxide (TCO). In general, solid-state dye sensitized solar cells (S-DSSCs) have a number of benefits over those based on a liquid electrolyte, particularly in terms of manufacturing, because printing electronics technologies may easily be adapted for their manufacture.

To prevent direct contact, a blocking layer must be present at the interface between the hole transporting material that transports holes and the SnO_2:F layer. By using the spray pyrolysis process, a compact TiO_2 layer has been used here between the nanocrystalline TiO_2 and SnO_2: F layer [35]. The compact layer's ideal thickness for our solid-state DSCs is roughly 100 nm. Then, a 2 µm thick nanoporous TiO_2 film layer was produced by doctor-blading TiO_2 colloidal paste on SnO_2:F conducting glass [16, 21]. After drying, the nanostructured TiO_2 film was burned at 450 °C for 30 minutes at a rate of 5 °C/min in an oven. The transparent TiO_2 films were submerged for two days in a RuL2 (NCS) 2: 2 TBA solution containing ruthenium dye. After drying under nitrogen flow at room temperature, the dye-coated nanocrystalline TiO_2 films were employed as photoanode electrodes. On top of the dye-coated TiO_2 films, thin films of P3HT spun-coated from chlorobenzene solution were applied to coat the thin hole transport layer. The hole carrying layers were typically between 100 - 150 nm thick. Following a second drying step, the top electrode, 40 nm gold (Au) film was vacuum-deposited on top of the hole-transporting layer. The entire devices were subjected to thermal annealing.

The photo-excited dye molecules sandwiched between the two semiconductors inject electrons into the TiO_2 conduction band and holes into the p-type polymer in the nanoporous TiO_2 semiconductor/Ru-dye/p-type polymeric semiconductor device combination. After the dye has injected electrons into the TiO_2, the hole-transporting polymeric material ought to be able to transfer holes from the dye. The holes are subsequently carried by hopping to the clear counter electrode. The counter electrode needs to be highly conductive, mechanically stable, and chemically and electrochemically stable. Several conditions must be met in order for this to be realized: (i) A method must be available for depositing a p-type polymeric material without dissolving or degrading the ruthenium dye monolayer on TiO_2 nano crystallites; (ii) the p-type polymeric material must be very transparent in the visible spectrum, where the ruthenium dye absorbs light; (iii) The dye's excited level (S*/S+) should be above TiO_2's conduction band (Ec), while the dye's ground state level (S0/S+) must be below the p-type polymeric material's (HOMO) upper edge of the valence band. This is necessary for both hole transfer to the valence band of the p-type polymeric material and electron transport from excited dye molecule to the conduction band of TiO_2; (iv) the upper edge of the valance band of p-type semiconductors (HOMO) must be placed above the ground state level of the dye (S0/S+) in order for the polymer hole conductor to be able to transport holes from the sensitized dye

after the dye has injected electrons into the TiO_2, as seen in Figure 4. (v) The transport properties of the p-type polymeric materials must be sufficient or high conductivity. Furthermore, previous studies have shown that the device performance can be enhanced by matching the energy level of each material [17-20]. The potential energy diagram of hybrid solid-state dye-sensitized solar cell components with respect to vacuum level is shown in Figure 4.

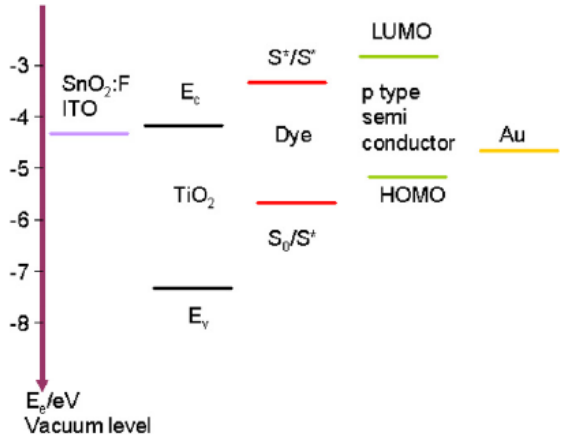
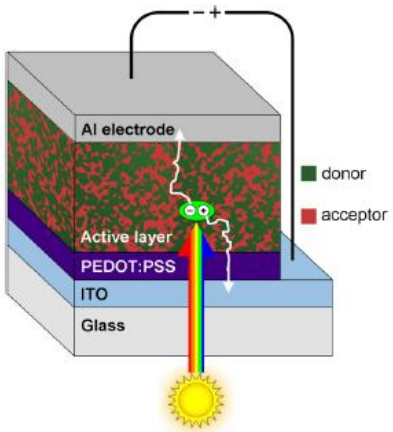


Figure 4. Energy diagram for an efficient charge transfer between hybrid solid-sate dye-sensitized solar cell components [19, 20].

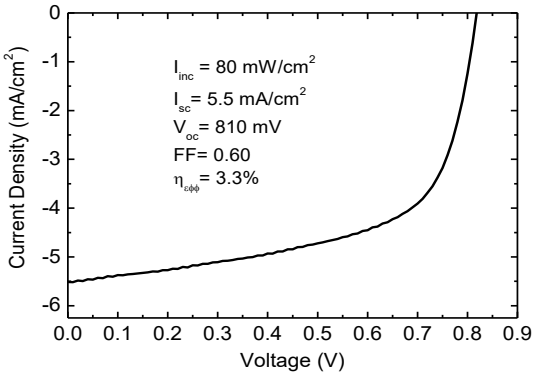
A solar simulator was used to light up photovoltaic devices. Using a calibrated single-crystal Si solar cell, the intensity of the incident light was measured. Keithley Source Meter measuring Unit (SMU) 2400 and Keithley SMU 236 measurements of current-voltage (I-V) characteristics were performed in a nitrogen atmosphere with light shining through the front of the photovoltaic devices. A Steuernagel solar simulator produced light (AM1.5 simulated irradiation at 80 and 100 mW/cm^2). The samples' temperature during measurement was 50 ^{0}C due to heat produced by the solar simulator.

3. RESULTS AND DISCUSSION

First, the device structure and characterization of bulk heterojunction ITO/PEDOT: PSS/MDMO-PPV: PCBM/LiF/Al organic polymer solar cells are presented in Figure 5 below.



(a)



(b)

Figure 5. Device structure of a bulk-heterojunction polymer-based organic photovoltaics (a) and current-voltage characteristics of annealed at 80 ^{0}C for 10 min, the best bulk-heterojunction polymer-based organic photovoltaics (b). The active layer in this type of device contains a donor and an acceptor organic material.

To investigate how thermal annealing affected the photovoltaics performance of PV devices of the number of devices tested for judging the reliability and accuracy, the bulk heterojunction MDMO-PPV: PCBM organic polymer solar cells at different annealing temperatures (60 ^{0}C, 80 ^{0}C, 100 ^{0}C, 120 ^{0}C) for 10 minutes were studied. This often involves systematically varying one parameter while keeping the other constant. Subsequently, the characteristic values of this annealed device at 80 ^{0}C for 10 minutes, the best device in Figure 5 (b) are V_{oc} = 810 mV, I_{sc} = 5.50 mA/cm^2 and a

fill factor FF = 0.60, and the resulting power conversion efficiency is 3.3% under standard AM1.5 illumination with an intensity of 80 mW/cm^2. Whereas the characteristics values of the unannealed bulk heterojunction MDMO-PPV: PCBM organic polymer solar cell are V_{oc} = 800 mV, I_{sc} = 4.75 mA/cm^2 and a fill factor FF = 0.58, and the resulting power conversion efficiency is about 2.8% under standard AM1.5 illumination (80 mW/cm^2)**.**

It is obvious that statistical analysis of a large sample size provides a more reliable and accurate representation of the device's performance. On the basis of several statistical and experimental analyses, the maximum power conversion efficiency of 3.3% for this device was found for a duration of the treatment of 10 minutes using an optimized device annealing temperature of 80 ^{0}C [Fig. 5 (b)]. This results demonstrated that the active layer's morphology is altered by thermal annealing of the device structure, which also alters the material's microstructure, crystallinity, phase separation structure and percolation routes, potentially improving device performance [3, 4, 6, 10, 36].This value is comparable to the previous report of power conversion efficiency of 3.5% obtained for P3HT: PCBM plastic solar cells after annealing for 5 minutes at 75 ^{0}C [36]. Additionally, using the P3HT:PCBM organic photovoltaic device, thermal annealing has shown to be a remarkable performance-enhancing step [37]. The above results demonstrated that the annealing resulted in an improvement photovoltaic properties and solar cell efficiency [14, 4, 3]. These enhancements seen in the short-circuit current density, fill factor and the power conversion efficiency can be explained by an increased charge carrier mobility due to a high degree of intermolecular order for both holes and electrons in the solution-caste photoactive layer and improved nanoscale morphology of the two components (donor/acceptor) in the photoactive layer. On the other hand, device architecture, such as inverted structures and the development of non-fullerene acceptors (NFAs) has led to significant improvements in PCE, with devices surpassing 19% efficiency [38]. Thus, this method helps create high-quality active layers with a more favorable molecular packing orientation, leading to better charge generation, separation, and transport.

Moreover, the photoactive layer morphology of the photovoltaic device is another critical factor for achieving high efficiencies, since the charge - separation and charge-transporting depend on the interface and the interpenetrated networks of a bulk heterojunction organic photovoltaic device [10, 4, 3]. In addition, the improvement of the device performance after annealing can be explained as follows: before annealing MDMO-PPV has an amorphous structure caused by the higher concentration of fullerene. During annealing PCBM diffuses out of the polymer matrix to form a percolated interpenetration of donor-acceptor network with improved electron transport properties. Consequently, the more ordered structure improves the contact between electrodes and photoactive film, which decreases the series resistance of the photovoltaic cells and increases the short circuit current as well as the fill factor. Obviously, there are a number of elements that might affect fill factor [39], but the main ones are charge carrier mobility (hole and electron) and mismatching in carrier mobility, charge collection due to interface state recombination, series and shunt resistances, and so on.

The device structure of an organic photovoltaic system based on bulk heterojunction polymers (a) and the current-voltage characteristics of an annealed bulk heterojunction polymer-based organic photovoltaic system (b) were displayed in Figure 5. Aluminum serves as the positive electrode whereas indium tin

oxide (ITO) is a popular transparent positive electrode. The diagram in Figure 5 (a) shows a bulk heterojunction active layer (MDMO-PPV: PCBM), where the blend of the donor and acceptor creates phase-segregated domains [39]. Interestingly several scientific works [5, 36, 41, 42, 43] were already published related to the relationship between the nanoscale morphology of the bulk heterojunction polymeric organic photoactive layer and the performance of the polymeric organic solar cells. Furthermore, previous reports [6, 5, 36, 40, 41, 42] attributed the improved photovoltaic performance to change in the bulk heterojunction organic materials nanoscale morphology induced by the donor-acceptor composition, the solvent and the thermal annealing of the organic photoactive layer. A recent complete study linking the morphology of the MDMO-PPV:PCBM mix made from chlorobenzene to solar cell performance was published [5, 41]. It is advised to compare the effects of reverse thermal annealing (RTA) and conventional/standard thermal annealing (TA) on the performance of organic photovoltaics (OPV). These studies can be shown in Table 1.

Table 1. Overview of comparative studies between conventional or standard thermal annealing (TA) and reverse TA (RTA) methods

Feature	Conventional Thermal Annealing (TA)	Reverse Thermal Annealing (RTA)
Primary goal	Improves phase segregation and molecular ordering for charge transport.	Enhances dielectric constant, forming a smaller Coulomb capture radius and optimal nano-scale aggregation and stacking orientation
Impact on film properties	Can improve vertical morphology, decrease defect density, and cause a de-doping process.	Affects the dielectric constant, which improves active layer quality and ordered molecular packing.
Typical outcome	Generally increases open-circuit voltage (Voc), short-circuit current density (Jsc) and fill factor (FF), leading to higher power conversion efficiency (PCE).	May lead to higher PCE by improving the active layer's charge carrier extraction and transport properties through enhanced film characteristics.

According to scholarly literatures in this study [3, 4, 6, 10, 36, 37, 43], the AFM/SEM images of the bulk heterojunction MDMO-PPV: PCBM organic polymer solar cells influenced by the donor-acceptor composition of materials, the solvent, device processing conditions and the thermal annealing of the organic photoactive layer demonstrated an enhancement to nano-scale morphology and phase separation. Hoppe *et al*. [44] also stated that, in accordance with theoretical predictions, experimental evidence indicates that the size of the phase-separated domain and the percolation of both hole- and electron-conducting phases are critical for enhancing the power conversion efficiency of organic solar cell devices. However, the morphological analyses (AFM/SEM/TEM images) of our PV devices were not investigated in this study due to a lack of necessary experimental equipment. A thorough investigation into the variation in energy level alignment brought on by thermal annealing and its associated impact on the Voc of organic solar cells has not yet been conducted, despite considerable efforts to explain the effect of thermal annealing on morphological alterations of solution-processed organic solar cells. Nonetheless, Ali et al. [37] showed that

thermal annealing is thought to help improve the Voc, Jsc, and FF in the polymer blend with improved charge transport characteristics for the devices that are annealed.

Second, the device structure and characterization of hybrid solid–state dye-sensitized photovoltaic device: SnO_2:F/compact TiO_2/ nc-TiO_2/Ru-dye/P3HT/Au is presented and shown in Figure 6. To investigate how thermal annealing affected the photovoltaics performance of PV devices, the hybrid solid–state dye-sensitized photovoltaic device: SnO_2:F/compact TiO_2/ nc-TiO_2/Ru-dye/P3HT/Au at different annealing temperatures (60 ^{0}C, 80 ^{0}C, 100 ^{0}C, 120 ^{0}C) for 5 minutes were studied. Consequently, the characteristic values of this annealed device at 100 ^{0}C for 5 minutes, the best device in Fig. 6 (b) are V_{oc} ~ 550 mV, I_{sc} ~ 11 mA/cm^2 and a fill factor FF ~ 0.53 under white light illumination with 100 mW/cm^2. The overall energy conversion efficiency, η_e for this annealed hybrid solid-state dye-sensitized solar cell was calculated to be over 3.2 % under standard AM1.5 illumination with intensity of 100 mW/cm^2. This maximum efficiency of the device was found for a duration of the treatment of 5 minutes for optimal annealing temperature at 100 ^{0}C.

Hence, by optimizing the morphology and reducing the density of defects, annealing enhances charge carrier mobility, leading to better short-circuit current density (Isc), open-circuit voltage (Voc) and fill factor (FF), which in turn boosts PCE. Whereas the characteristics values of the unannealed solid–state hybrid photovoltaic device: SnO_2: F/compact TiO_2/ nc-TiO_2/Ru-dye/P3HT/Au are V_{oc} ~ 530 mV, I_{sc} ~ 9.54 mA/cm^2 and a fill factor FF ~ 0.48, and the resulting power conversion efficiency is 2.4% under standard AM1.5 illumination (100 mW/cm^2). The rise in device performance following thermal annealing can be attributed to the burning of the shunts. The increase in short-circuit current density suggests an enhancement in the mobility of the charge carriers inside the photoactive layer as well as the filling of transparent hole conductors into the pores between the TiO_2 nanoparticles. Previous studies [16, 17, 18, 19] demonstrated that annealing helps achieve a stable, well-ordered morphology in the active layer, including appropriate phase separation and crystallinity, which are vital for efficient charge separation and transport. Novel p-type semiconductor materials and various techniques for depositing p-type material onto dye-coated nanoporous films are also likely to increase the efficiency of this cell. Figure 6 below demonstrated the (a) Device structure [20] and (b) I-V characteristics of annealed hybrid solid-state dye-sensitized photovoltaic devices.

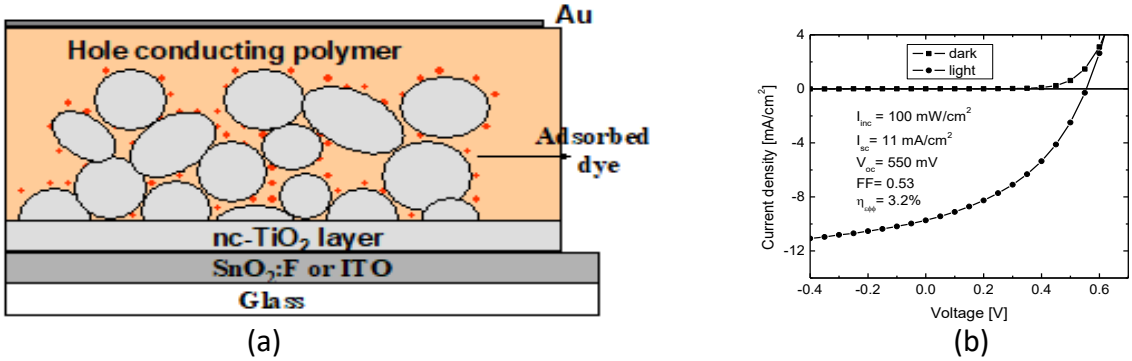


Figure 6. (a) Device structure and (b) I-V characteristics of annealed at 100 ^{0}C for 5 min, the best hybrid solid-state dye-sensitized photovoltaic devices: SnO_2:F/compact TiO_2/nc-TiO_2/Ru-dye/P3HT/Au (solid squares: dark, solid circles: illuminated with 100 mW/cm^2).

According to the scholarly literatures in this study [16, 17, 18, 19, 20, 21], the AFM/SEM images of the hybrid solid–state dye-sensitized photovoltaic devices influenced by thermal annealing of the hybrid solid-state dye-sensitized photovoltaic devices demonstrated an enhancement to nano-scale morphology and infiltration of the organic / polymer hole-conductor into nanoporous TiO_2. However, the morphological analyses (AFM/SEM/TEM images) of our PV devices were not investigated in this study due to a lack of necessary experimental equipment.

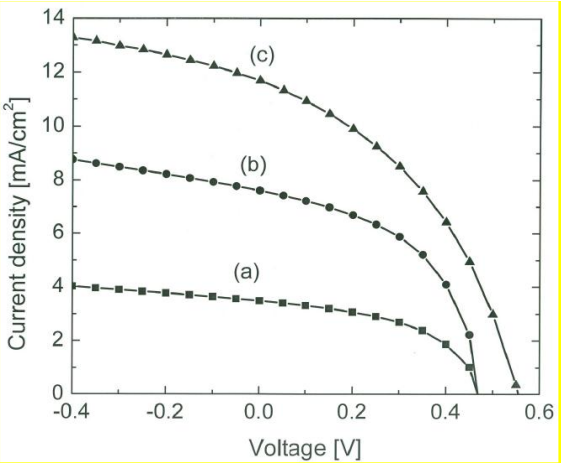


Figure 7 Current-voltage characteristics of solid-state DSCs upon three different ways of infiltration of the polymer hole-conductor P3HT into nanoporous TiO_2 under white light illumination with 100 mW/cm^2: (a) 10 times repeated spin-coating of the P3HT solution, (b) soaking in the solvent with subsequent spin-coating of the P3HT solution, and (c) soaking in the solvent with subsequent spin-coating of the P3HT solution and annealing treatment at 100 ^{0}C for 5 minutes

Third, Figure 7 shows the measured I-V curves and photovoltaic performance metrics of solid-state hybrid DSCs: SnO_2:F/nc-TiO_2/Ru-dye/P3HT/Au under white light illumination with 100 mW/cm^2.

It was discovered that the effectiveness of solid-state dye-sensitized solar cells depends on how well Titania nanostructures are filled with a solid hole-conducting material. Since a transparent hole conductor was not completely filled into the pores between the nanoparticles of TiO_2 and because fewer charges were removed, solution casting, or one-time spin coating, of the polymer, performed poorly. The polymer hole-conductor P3HT, however, infiltrated into Titania nanostructures in three different ways, each of which resulted in an effective photovoltaic performance. Filling by ten times repeated spin coating granted reasonable efficiencies up to 0.91% [Fig. 7 (a)]. However, soaking nanoparticles of TiO_2 film in toluene solvent with subsequent spin coating of the P3HT solution showed higher efficiencies up to 2% [Fig.7 (b)], and soaking in toluene solvent with subsequent spin coating of the P3HT solution as well as annealing treatment at 100 ^{0}C for 5 minutes resulted in the best efficiencies up to 3.6% % [Fig.7 (c)] under white light illumination with 100 mW/cm^2. Hence, efficient solid-state hybrid DSCs performances demonstrated that an increasing filling of the pores in the Titania nanostructures with P3HT is observed from filling by repeated spin coating via filling, by solution casting to filling by soaking in the solvent with subsequent spin coating and annealing treatment at 100 ^{0}C for 5 minutes. The results show the usefulness of conducting polymers as hole-transport materials for the possible replacement of the liquid electrolyte in nanoporous TiO_2 type solar cells and optimization of these devices is a promising approach from both academic and industrial sides.

Table 2 shows the measurements of summarized photovoltaic parameters of solid-state DSCs for the different kinds of P3HT infiltration into dye-coated TiO_2 nanostructures. These results indicate that the performance of solid-state DSCs can be strongly affected by the device processing conditions. The improved short-circuit current density (I_{sc}) as high as 12 mA/cm^2 in Table 2 is mainly due to the improved charge carrier mobilities and ascribed to the enhanced charge transport rate associated with the improved charge collection efficiency. Additionally, we

observe that open circuit voltage (V_{oc}) has slightly higher value upon annealing which can be surmised due to reduced charge recombination pathways because of a different local morphology in the polymer for the annealed device.

Table 2. Overview of solid-state DSCs parameters for the different kinds of P3HT infiltration into dye-coated TiO_2 nanostructures

Way of Filling	Voc [mV]	Isc [mA/cm^2]	FF	η[%]
10 times spin-coating of the P3HT solution	450	3.5	0.58	0.91
Soaking in the solvent with subsequent spin coating of the P3HT solution	450	7.6	0.6	2
Soaking in toluene solvent with subsequent spin coating of the P3HT solution as well as annealing treatment at 100 ^{0}C for 5 minutes	550	12	0.54	3.6

This maximum power conversion efficiency η_e of the solid-state hybrid DSC device was found due to an enhancement of the mobility of the charge carriers inside the photoactive layer and upon filling of nanoporous TiO_2 film by soaking in the solvent with subsequent spin-coating of hole conducting polymer material, P3HT as well as for a duration of the treatment of 5 minutes in nitrogen atmosphere for annealing the solid-state DSC using a temperature of 100 ^{0}C. Additionally, the photo-physical study conducted by Sai Santosh Kumar et al. [45] revealed that the increased crystallinity of the P3HT thin-film after thermal annealing results in increased hole-transfer from the organic dye to polymer, along with increased hole-mobility and decreased charge recombination. These factors directly influence the observed improvement in the device performances.

Finally, the photovoltaic external quantum efficiency (η_c) or the spectrally resolved incident photon-to-converted electron efficiency (IPCE) is defined as the number of electrons flowing through an external circuit under short circuit conditions per incident photon should be mentioned and is calculated from the spectrally resolved short-circuit current,

$$\eta_c(\%) = 1240/\lambda[nm] x I_{sc}[\mu A/cm^2]/I_{inc}[W/m^2] \quad (3)$$

Where I_{inc} is the intensity of the incident light. Figure 8 displayed a comparison between the spectrum resolved photon-to-electron conversion efficiency (IPCE) of a solid-state hybrid dye-sensitized nanocrystalline TiO_2 solar cell under xenon arc lamp illumination from the TiO_2 side plotted on the right axis and the optical density on the left axis of a solid–state hybrid dye-sensitized nc-TiO_2 solar cell. One can see very clearly, that the two curves have the same onset and spectrally resolved maxima at 536 nm. The maximum value of the photon-to-electron conversion efficiency of a solid–state hybrid dye-sensitized nc-TiO_2 solar cell was calculated to be about 72% under monochromic illumination at 536 nm with power intensity of 0.1mW/cm^2.

Furthermore, the decline of the IPCE above 600 nm towards the red is caused by the decrease in the extinction coefficient of RuL2(NCS)2. This enhancement of the IPCE originates from an enhanced absorption, electron transfer and transport due to an enhancement of the charge carrier mobility of the bulk.

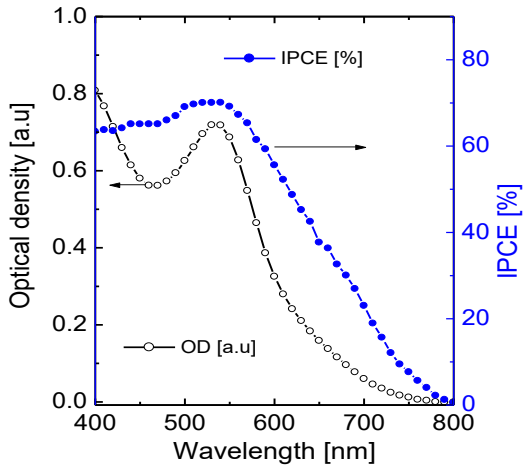


Figure 8. Spectral resolved photon-to-electron conversion efficiency, IPCE (solid circles) of a solid-state hybrid dye-sensitized nanocrystalline TiO_2 solar cell (plotted on the right axis) through illumination from TiO_2 side by xenon arc lamp and optical absorption (open circles) for nanocrystalline TiO_2 film coated with ruthenium-dye (plotted on the left axis).

4. CONCLUSIONS

In conclusion, organic and hybrid solid–state dye-sensitized solar cells are considered as emerging third generation technologies to replace conventional silicon solar cells as well as thin film solar cells. Experimentally, in optimized annealed devices, we achieved above 3% power conversion efficiencies performances under AM 1.5 global illuminations at 80 mW/cm^2 as well as 100 mW/cm^2. In addition, we observed an incident photon to collected electron efficiency (IPCE) of 72%. These advances are a testimony to the rapid progress being made in the field of organic and dye-sensitized emerging 3rd generation photovoltaic technologies. The impact of thermal annealing on the performance of organic and solid-state dye-sensitized photovoltaics devices, as well as a more comprehensive contextual discussion regarding morphological optimization, device performance enhancements, and a more efficient and economical production process in comparison to the current body of knowledge, would therefore be crucial. The conversion of sunlight to electricity by organic and hybrid solid-state dye-sensitized solar cells is very interesting and promising since organic and dye-sensitized solar cells offer the possibility of fabricating large area, low-temperature processing, flexible, lightweight, cost-effective devices using simple and environmentally friendly techniques. Obviously, roll-to-roll processing is at the heart of organic photovoltaics as well as solid-state dye-sensitized solar cells in the future and that the successful realization of lowcost emerging 3rd generation photovoltaics will be closely linked to this. These advantages have so far drawn a lot of research attention and commercialization effort. Furthermore, overcoming limitations in device stability, particularly "burn-in" degradation and long-term operational stability, remains a key challenge and area of future research.

CONFLICT OF INTEREST

The author affirms that there were no financial or commercial ties that might be interpreted as a potential conflict of interest during the research.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the Abdus Salam International Centre for Theoretical Physics (ICTP)-Trieste, Italy for financial support during his research visit study as a Regular Research Associate of ICTP and Department of Physics/DTE, CELS, Addis Ababa University for support in any kind.

REFERENCES

[1]. M. Green, E. Dunlop, J. Hohl-Ebinger, M. Yoshita, N. Kopidakis, X. Hao, "Solar cell efficiency tables (version 57)," *Prog. Photovolt. Res. Appl.* Vol. 29, no.1, p. 3-15, 2020.

[2]. E. K. Solak, and E. Irmak, "Advances in organic photovoltaic cells: a comprehensive review of materials, technologies, and performance," *RSC Adv*. Vol. 13, p. 12244–12269, 2023.

[3]. W. Aloui, T. Adhikari, J-M. Nunzi, A. Bouazizi, K. Khirouni, "Effect of thermal annealing on the electrical properties of P3HT:PC70BM nanocomposites," *Materials Science in Semiconductor Processing*, vol. 39, p. 575–581, 2015.

[4]. M. K. Hamad, H. F. Al-luaiby, W. A. Hussain, A. K. Hassan, "Post-fabrication annealing effects on the performance of polymer: Fullerene solar cells with ZnO nanoparticles," *Chemistry and Materials Research*, vol. 6, no. 6, p. 15-22, 2014.

[5]. S. E. Shaheen, C. J. Brabec, N. S. Sariciftci, F. Padinger, T. Fromherz, and J. C. Hummelen, "2.5% efficient organic plastic solar cells," *Appl. Phys. Lett*. vol.78, no. 6, p. 841-843, 2001.

[6]. D. Gebeyehu, C. J. Brabec, F. Padinger, T. Fromherz, J. C. Hummelen, D. Badt, H. Schindler, N. S. Sariciftci, "The interplay of efficiency and morphology in photovoltaic devices based on interpenetrating networks of poly(3-octylthiophene) with fullerenes," *J. Synth. Met*., vol. 118, p. 1-9, 2001.

[7]. G. Yu, J. Gao, J. C. Hummelen, F. Wudl, A. J. Heeger, "Polymer photovoltaic cells: Enhanced efficiencies via a network of internal donor-acceptor heterojunctions," *Science* vol. 270, no. 5243, p. 1789-1791, 1995.

[8]. P. Peumans, A. Ykimov, S. R. Forrest, "Small molecular weight organic thin-film photo- detectors and solar cells," *J. Appl. Phys.,* vol. 93, no.7, p. 3693-3723, 2003.

[9]. D. Gebeyehu, B. Maennig, J. Drechsel, K. Leo, M. Pfeiffer, "Bulk heterojunction photovoltaic devices based on donor-acceptor organic small molecule blends, *Solar Energy Materials and Solar Cells,*" vol. 79, no. 1, p. 81-92, 2003.

[10]. Z. Yi, W. Ni, Q. Zhang, M. Li, B. Kan, X. Wan, and Y. Chen, "Effect of thermal annealing on active layer morphology and performance for small molecule bulk heterojunction organic solar cells," *J. Mater. Chem. C*, vol. 2, p. 7247-7255, 2014.

[11]. B. O'Regan, M. Grätzel, "A low-cost, high efficiency solar cells based on dye-sensitized colloidal TiO_2 films. *Nature,* vol, 353, p. 737-740, 1991.

[12]. R. Knödler, J. Sopka, F. Harbach, H. W. Grünling, "Photoelectrochemical solar cells based on dye-sensitized colloidal TiO_2 layers," *Solar Energy Materials and Solar Cells*, vol. 30, no. 3, p. 277-281. 1993.

[13]. C. J. Barbé, F. Arendse, P. Comte, M. Jirousek, F. Lenzmann, V. Shklover, M. Grätzel, "Nanocrystalline TiO_2 electrodes for photovoltaic applications," *J. Am. Ceram. Soc.* vol. 80, p 3157-3171. 1997.

[14]. N.-G. Park, G. Schlichthörl, J. van de Lagemaat, H. M. Cheong, A. Mascarenhas, A. J. Frank, "Dye-sensitized TiO_2 solar cells: Structural and Photoelectrochemical Characterization of Nanocrystalline Electrodes Formed from the Hydrolysis of $TiCl_4$," *J. Phys. Chem. B*. vol. 103, no. 17, p. 3308-3314, 1999.

[15]. D. Gebeyehu, and N. S. Sariciftci, "Photoelectrochemical Solar Cells Based on Dye-Sensitization of Nanocrystalline TiO_2," *Ethiopian Journal of Engineers and Architectures,* vol. 25. p. 63-72, 2008.

[16]. U. Bach, D. Lupo, P. Comte, J. E. Moser, F. Weissoertel, J. Salbeck, H. Spreitzer, M. Graetzel, "Solid-state dye-sensitized mesoporous TiO_2 solar cells with high photon to electron conversion efficiencies," *Nature,* vol. 395, p. 583-585, 1998.

[17]. S. Günes, & N. S. Sariciftci, "Hybrid solar cells," *Inorganica Chimica Acta,* vol. 361, no. 3, p. 581-588, 2008.

[18]. U. Mehmood, A. Al-Ahmed, F. A. Al-Sulaiman, M. I. Malik, F. Shehzad, A. U. H. Khan, "Effect of temperature on the photovoltaic performance and stability of solid-state dye-sensitized solar cells: A review," Renewable and Sustainable Energy Reviews.' vol. 79, p. 946-959, 2017.

[19]. N. Vlachopoulos, M. Gratzel, A. Hagfeldt, "Solid-state dye-sensitized solar cells using polymeric hole conductors,' *RSC dv.,* vol. 11, no. 62, p. 39570–39581.

[20]. D. Gebeyehu, J. C. Brabec, N. S. Sariciftci, D. Vangeneugden, R. Kiebooms, D. Vanderzande, F. Kienberger, H. Schindler, "Hybrid solar cells based on dye-sensitized nanoporous TiO_2 electrodes and conjugated polymers as hole transport materials." *Synth. Met.* Vol. 125, p. 279-287, 2002.

[21]. J-K. Ding, J. Melas-Kyriazi, N-L. Cevey-Ha, K. G. Chittibabu, S. M. Zakeeruddin, M. Graetel., M. D. McGehee, "Deposition of hole-transport materials in solid-state dye-sensitized solar cells by doctor-blading," *Organic Electronics*, vol. 11, p.1217-1222, 2010. [22]. C. Girotto, and J. Genoe, "Spray coating for fabricating polymer-based organic solar cells," *Global Solar Technology. International Edition*, vol. 3, no. 3, p. 10-13, 2010.

[23]. M. A. Green, K. Emery, Y. Hishikawa, W. Warta, "Solar cell efficiency tables (version 37)," *Prog. Photovolt: Res. Appl*,. vol. 19, p. 84–92, 2011.

[24]. D. R. Gamota, P. Brazis, K. Kalyanasundaram, J. Zhang, Printed Organic and Molecular Electronics, Springer, 2004, ISBN 978-1402077074.

[25]. C. Brabec, U. Scherf, V. Dyakonov, Organic Photovoltaics: Materials, Device Physics, and Manufacturing Technologies, Wiley-VCH, 2008, ISBN 978-3527316755.

[26]. F. C. Krebs, "Polymer solar cell modules prepared using roll-to-roll methods: Knife-over edge coating, slot-die coating and screen printing," *Solar Energy Materials and Solar Cells,* vol. 93, no. 4, p. 465-475, 2009.

[27]. G. Schwarz, B. Maennig, C. Ulrich, W. Gnehr, S. Sonntag, Q. Erfurth, E. Wollrab, K. Walzer, M. Pfeiffer, "Efficient and long-term stable organic vacuum deposited tandem solar cells," *Proc. of SPIE* 7446**,** Organic Photovoltaics X, 2009, 74160K.

[28]. G. Griffini, J. D. Douglas, C. Piliego, T. W. Holcombe, S. Turri, J. M. J. Fréchet, J. L. Mynar, "Long-term thermal stability of high-efficiency polymer solar cells based on photocrosslinkable donor-acceptor conjugated polymers," *Advanced materials*, vol. 23, no. 14, p. 1660-1664, 2011.

[29]. G. Chidichimo, and F. Filippelli, "Organic Solar Cells: Problems and Perspectives," *International Journal of Photoenergy,* Article ID 123534, p. 1-11, 2010.

[30]. S. N. F. Mohd-Nasir, M. Y. Sulaiman, N. Ahmad-Ludin, M. A. Ibrahim, K. Sopian, and M. A. Mat-Teridi, **"**Review of Polymer, Dye-Sensitized, and Hybrid Solar Cells," *International Journal of Photoenergy,* p. 1-12, 2014.

[31]. D. Khatiwada, S. Venkatesan, J. Chen, Q. Chen, N. Adikari, A. Dubey, A. F. Mitul, L. Mohammad, J. Sun, C. Zhang, L. Luo, Q. Qiao, "Morphological Evolution and its Impacts on Performance of Polymer

Solar Cells." *IEEE Transactions on Electron Devices*, vol. 62, no. 4, p. 1284-1290, 2015.

[32]. F. C. Krebs, "Fabrication and processing of polymer solar cells: A review of printing and coating techniques," *Solar Energy Materials and Solar Cells*, vol. 93, no. 4, p. 394-412, 2009.

[33]. R. Søndergaard, M. Hösel, D. Angmo, T. T. Larsen-Olsen, and F. C. Krebs, "Roll-to-roll fabrication of polymer solar cells," *Materials today*, vol. 15, no. 1-2, p. 36-49, 2012.

[34]. L. Travis, Benanti & D. Venkataraman, "Organic solar cells: An overview focusing on active layer morphology," *Photosynthesis Research*, vol. 87, p. 73–81, 2006.

[35]. L. Kavan, and M. Grätzel, "Highly efficient semiconducting TiO_2 photoelectrodes prepared by aerosol pyrolysis," *Electrochimica Acta*, vol. 40, no. 5, p. 643-652, 1995.

[36]. F. Padinger, R. S. Rittberger, and N. S. Sariciftci, "Effects of postproduction treatment on plastic solar cell," *Adv. Funct. Mater.* vol. 13, no. 1, p. 85-88, 2003.

[37] Y. Ali, N. P. Holmes, N. Cooling, J. Holdsworth, W. Belcher, P. Paul Dastoor, X. Zhou, "Optimization of bulk heterojunction organic photovoltaics," *Coatings*, vol. 13, no. 1293, p. 1-12, 2023.

[38]. W. Yufei et al. "A novel upside-down thermal annealing method toward high quality active layers enables organic solar cells with efficiency approaching 20%", Advanced Materials, vol. 36, no. 47, p. 1-11, 2024.

[39]. G. F. A. Dibb, F. C. Jamieson, A. Maurano, J. Nelson, J. R. Durrant, "Limits on the fill factor in organic photovoltaics: Distinguishing non-geminate and geminate recombination mechanisms," *Journal of Physical Chemistry Letters*, vol. 4, p. 803-808, 2013.

[40]. G. Li, R. Zhu, Y. Yang, "Polymer solar cells," *Nature Photonics,* vol. 6, p.153-161, 2012.

[41]. J. K. J. Van Duren, X. Yang, J. Loos, C. W. T. Bulle-Lieuwma, A. B. Sieval, J. C. Hummelen, and R. A. J. Janssen, "Relating the morphology of poly (p-phenylene vinylene)/methanofullerene blends to solar cell performance," *Adv. Funct. Mater*. vol. 14, no. 5, p. 425-434, 2004.

[42]. Y. Kim, S. A. Choulis, J. Nelson, D. D. C. Bradley, S. Cook, J. R. Durrant, "Device annealing effect in organic solar cells with blends of regioregular poly(3-hexylthiophene) and soluble fullerene," *Appl. Phys. Lett*., vol. 86, no. 6, p. 063502-063502-3, 2005.

[43]. W. L. Ma, C. Y. Yang, X. Gong, K. Lee, and A. J. Heeger, "Thermally stable efficient polymer solar cells with nanoscale control of the interpenetrating network morphology," *Advanced Functional Materials*, vol. 15, no. 10, p. 1617-1622, 2005.

[44]. H. Hoppe, T. Glatze, M. Niggemann, W. Schwinger, F. Schaeffler, A. Hinsch, M. Ch. Lux-Steiner, N. S. Sariciftci, "Efficiency limiting morphological factors of MDMO-PPV: PCBM plastic solar cells," *Thin Solid Films*, vol. 511-512, p. 587-592, 2006.

[45]. R. Sai Santosh Kumar, R. Grancini, A. Petrozza, A. Abrusci, H. J. Snaith, and G. Lanzani, "Effect of polymer morphology on P3HT-based solid-state dye sensitized solar cells: an ultrafast spectroscopic investigation," *OPTICS EXPRESS*, A470 21 (S3), 2013.