Open Access Research Journal of
Engineering and Technology
Journals home page: https://oarjpublication/journals/oarjet/
ISSN: 2783-0128 (Online)

(Research Article)

MLeNet -CNN fusion model for crop yield prediction with extended normalization process

Siddhant Hanumant Kale [1], Shailesh Deshmukh [2] and D. S. Shelar [3,*]

[1] Research Scholar, Bir Tikendrajit University, Manipur, India and Lecturer, AISSMS Polytechnic, Pune, India.
[2] Associate Professor, Bir Tikendrajit University, Manipur, India.
[3] Assistant Professor, Department of Engineering Sciences, AISSMS Institute of Information Technology, Pune, India.

Open Access Research Journal of Engineering and Technology, 2025, 09(02), 084-102

Publication history: Received on 04 November 2025; revised on 10 December 2025; accepted on 12 December 2025

Article DOI: https://doi.org/10.53022/oarjet.2025.9.2.0100

Abstract

Crop yield prediction (CYP) at the field level plays a crucial role in evaluating agricultural commodities plans for import-export strategies, agricultural production and increasing farmer incomes. Traditional methods, which depend on historical data, weather conditions, and agronomic models, have been greatly improved through the application of machine learning technologies. This paper proposes an innovative M-Let and CNN-based Fusion model for Crop Yield Prediction (MCF-CYP) model. The input data undergoes preprocessing, where a Three-tier Data Normalization (TDN) technique is applied to standardize the values, ensuring all features are on a consistent scale for improved model effectiveness. Next, feature extraction is performed to identify the key characteristics of the data. In this case, features like Beckenstein Hawking with Deng Belief-based Renya Entropy (BHDB-RE) features, information gain, and various statistical measures are extracted to capture essential insights related to crop yield. These features are then passed into the prediction model, which utilizes two deep learning architectures: Multi-head LeNet (M-Let) and Convolutional Neural Networks (CNN). Both models analyze the complex patterns in the data and generate accurate crop yield predictions. Finally, the predicted yield is outputted, providing valuable insights for crop management and other agricultural decisions.

Keywords: Crop Yield Prediction; Deep Learning; Renya Entropy; Three-Tier Data Normalization

1. Introduction

Agriculture plays a central role in the global economy, providing raw materials and food for industries and supporting livelihoods in many regions of the world. Farming involves growing crops and raising animals, which supply essential resources such as food and other raw materials like cotton, wool, leather, and paper products [1]. Accurate and timely crop yield estimation offers substantial benefits to farmers, agribusinesses and food producers, by enabling better planning for harvesting, marketing, labor allocation and resource management [2]. Reliable crop yield forecasts also facilitate decision-making for policymakers, which helping them address challenges related to food security and agricultural sustainability [3]. Traditional methods of yield estimation, such as prior knowledge and field surveys of growing conditions, have limitations. However, advancements in data-driven technologies, particularly machine learning (ML) and remote sensing, have significantly improved the accuracy and efficiency of CYP [4].

Machine learning, especially in combination with remote sensing data, has gained increasing prominence in agricultural research [5]. ML techniques, such as deep neural networks, convolutional neural networks, and recurrent neural networks, have demonstrated superior accuracy in predicting crop yields across various agricultural systems [6]. These models are adept at handling complex, non-linear relationships between input variables like weather conditions, soil properties, and crop management practices, which are critical to understanding and predicting crop performance [7].

* Corresponding author: D. S. Shelar

Unlike conventional regression models, ML models are non-parametric, capable of fitting complex functions and handling large datasets, which makes them well-suited for the challenges posed by modern agriculture [8].

However, despite the advances in ML for crop yield prediction, challenges remain. Climate change, which induces increasingly extreme weather conditions, can significantly affect crop growth and yield [9]. These climate variables often fall at the extremes of the weather distribution, making prediction more difficult [10]. Moreover, limited access to accurate and high-quality data in many regions poses another hurdle. The need for reliable, high-resolution datasets remains a challenge in parts of the world where data collection infrastructure is underdeveloped [11].

In recent years, the combination of machine learning and remote sensing has provided valuable insights into crop performance and potential yield outcomes [12]. Satellite imagery, for example, offers detailed spatial and temporal information about crop conditions, enabling the prediction of yields with greater accuracy [13]. Precision agriculture, which incorporates such technologies, aims at optimizing farming practices based on within-field variability [14]. This approach offers potential solutions for improving yield predictions, enhancing agricultural management, and addressing the increasing demands of food production [15].

Despite these advancements, the reliability and interpretability of CYP models remain crucial [16]. As ML models become more complex, ensuring that predictions are understandable to farmers, policymakers, and other stakeholders is essential for building trust and facilitating decision-making [17]. Therefore, improving model accuracy, efficiency, and interpretability is an ongoing challenge that requires careful consideration of both technical and practical aspects of crop yield prediction. This paper proposes an innovative M-LNet and CNN –based Fusion model for Crop Yield Prediction (MCF-CYP) model. The key contributions of this study are as follows:

Adopting a three-tier data normalization technique for preprocessing the input data that considers the interdependencies among the feature of each vector and the absolute value of features.

Extracting Bekenstein-Hawking with Deng Belief-based Renyi entropy features in the feature extraction step. This can capture complex relationship in the data and the parameter tune the entropy metrics based on the characteristics of data.

Proposing M-LNet and CNN–based Fusion model for crop prediction, in which the M-LNet model enhances the stability of training and faster convergence.

The rest of the paper is organized as follows: Section 2 provides a review of related work, highlighting previous research and methodologies in crop yield prediction using deep learning models. Section 3 outlines the architecture of the proposed model, detailing the CNNs and M-LNet architecture used for crop yield prediction. Section 4 discusses the results of the model's performance, comparing it against traditional methods and other deep learning approaches. Finally, Section 5 concludes the paper.

2. Literature Review

In 2024, S. M. Mirhoseini Nejad et al., [18] introduced a soybean yield prediction model combining ConvLSTM, 3D-CNN, and ViT, using multispectral remote sensing data. The model captures both spatial and temporal patterns, significantly improving prediction accuracy with lower RMSE and higher correlation coefficients. This approach enhances decision-making in accurate agriculture, aiding in resource optimization, irrigation, and fertilizer application. The model is robust across different conditions, applicable to various crops and regions, and provides actionable insights for agricultural stakeholders.

In 2023, Uferah Shafi et al., [19] presented a framework for predicting wheat grain yield using three regression techniques like RF, XGB, and LASSO. The models are evaluated with the aid of drone-based multispectral data collected from three wheat fields such as SD1, SD2, SD3. The analysis showed that LASSO outperforms the other models, achieving a coefficient of determination and a MAE effectively. The predicted annual yields were determined for SD2, SD1, and SD3. This framework helps agronomists and farmers make informed decisions regarding crop management, including harvest plans planting schedules and resource allocation.

In 2024, Md Jiabul Hoque et al., [20] introduced a crop yield prediction system that integrates a year’s worth of meteorological data, pesticide records and crop yield data, using machine learning techniques. Three models, named as GB, KNN, and MLR that were trained and evaluated with rigorous data preprocessing and hyperparameter tuning through K-Fold cross-validation and Grid Search CV to prevent overfitting. Moreover, the Gradient Boosting model achieved an impressive high R^2, highlighting its accuracy in predicting crop yields.

In 2024, A. Badshah et al., [21] presented two ML models for agricultural applications such as one for crop recommendation and another for wheat yield prediction. For crop recommendation, a dataset from Kaggle, including soil pH, humidity, temperature and nutrient levels, was analyzed using classification schemes like LR, DT, RF, KNN, GNB, Extra Tree Classifier and SVM. RF achieved the highest performance with high accuracy. The SVR model, after hyperparameter tuning, achieved high accuracy for wheat yield prediction. These approaches can help boost agricultural productivity, reduce risks, improve food security and encourage sustainable farming practices.

In 2024, Alakananda Mitra et al., [22] utilized ML to predict cotton yield, considering factors like climate change, soil diversity, cultivars, and fertilizer use. Data from the GOSSYM model, were analyzed across nine locations in Mississippi, Texas, and Georgia. The study used gathered heat as a substitute for time-series weather data to reduce computational complexity. Several models including SVR, Light GBM and RF were computed.

In 2023, S. M. M. Nejad et al., [23] proposed two ML architectures for improving crop yield prediction accuracy. The first model combined skip connections, 2D-CNN and LSTM-Attentions, while the second model integrated skip connections, 3D-CNN and Conv LSTM Attention. Using MODIS data for soybean cultivation, the second model outperformed other techniques.

In 2022, S. P. Raja et al., [24] highlights the importance of efficient feature selection for improving the accuracy of ML models in crop yield prediction. By eliminating irrelevant or redundant features, the model becomes more computationally efficient and precise. Optimal feature selection ensures that only the most relevant data was used, preventing unnecessary complexity and improving model performance.

In 2024, Mohammad Amin Razavi et al., [25] uses advanced data-driven techniques to explore the relationships between crop yields and geographical features in Senegal, utilizing remotely sensed data. The authors applied apply correlation analysis and clustering to uncover important patterns as well as optimize model performance namely XG Boost, Cat Boost, RF, and Light GBM. Cat Boost and XG Boost outperformed the others, and generated synthetic crop data by Variational Auto Encoder helped to address dataset limitations, enhancing model robustness and reducing overfitting. The approach improves predictions and

2.1. provides valuable insights for policymakers and farmers in data-scarce regions

In 2024, Leelavathi Kandasamy Subramaniam and Rajasenathipathi Marimuth [26] presented an efficient approach for CYP of Indian crops using DL and DR. The process included data preprocessing (cleaning and normalization), dimensionality reduction with SEKPCA, and classification using a WTDCNN to predict high crop yield profit.

In 2024, Che-Hao Chang et al., [27] proposed a hybrid deep learning model for rice CYP, combining classification and regression models through shared layers. Statistical analysed such as PCC, SHAP, and RFECV select relevant features to reduce training time. Data preprocessing normalized and reshaped the data into a three-dimensional matrix.

Further, these above works are more descriptively tabulated in Table I by pinpointing those features and challenges of these works.

Table 1 Features and challenges of Conventional works

Author [Citation]	Methodology	Features	Challenges
S. M. Amirhossein Nejad et al., [18]	3D-CNN, Conv LSTM, and Vitt	Improves sustainable farming practices by optimizing resources and reducing waste	High computational requirements due to the model's complexity and large number of parameters
Ferah Shafi et al., [19]	XGB regression, RF, and LASSO regression	It utilizes multispectral data for a more detailed and accurate analysis of crop health	Lack of integration with soil and climate data which could further enhance model accuracy
Md Jiabul Hoque et al., [20]	GB, MLR and KNN	Highly generalizable models capable of handling large-scale datasets effectively	The model's accuracy heavily relies on data quality, and inaccurate data could affect predictions
A. Badshah et al., [21]	RF, SVM, and SVR	Transparency and trust in model outputs through XAI, improving farmer decision-making	It requires large, high-quality datasets, which may not always be available for all regions or crops
Alakananda Mitra et al., [22]	ML	Reduced computational complexity by converting time-series weather data into scalar values	It may struggle with data outside the training space, requiring retraining with new data
S. M. M. Nejad et al., [23]	2D-CNN, 3D-CNN, LSTM-Attention, Conv LSTM	Attention mechanism improves model focus on main features, increasing efficiency and accuracy	It does not incorporate additional climatic factors like watering and genotype, which could improve accuracy.
S. P. Raja et al., [24]	ML	Feature selection process reduces redundancies and improves model efficiency	Model complexity could increase with large datasets, impacting computational efficiency
Mohammad Amin Razavi et al., [25]	ML	Improved predictive performance through hyperparameter tuning and synthetic data augmentation	It requires ongoing monitoring (MLOPS) for long-term functionality and performance
Leelavathi Kandasamy Subramaniam and Rajasenathipathi Mari Muth [26]	WTDCNN	It addressing saturation of vanishing gradient problem and minimizing the network's prediction loss.	The methodology proposed can seamlessly integrate with precision agriculture tools, enabling real-time CYP and aiding farmers in making informed decisions.
Che-Hao Chang et al., [27]	Multitasking method	The use of a multitasking method boosts regression accuracy and reduces the risk of overfitting.	Need to explore how generative adversarial networks can be utilized to supplement real data, improving the performance of the proposed hybrid model.

2.2. Problem Definition

Based on the literature review, it is evident that while ML techniques have shown promise in improving crop yield prediction and agricultural decision-making, several challenges persist in their widespread application. In existing works, preprocessing was often conducted in S. M. Amirhossein Nejad et al., [18], Ferah Shafi et al., [19], Md Jiabul Hoque et al., [20], S. P. Raja et al., [24], Leelavathi Kandasamy Subramaniam and Rajasenathipathi Mari Muth [26] and Leelavathi Kandasamy Subramaniam and Rajasenathipathi Mari Muth [26]. To enhance the interdependencies between the absolute values of features and the features of each vector, enhanced data normalization technique is required. For predicting crop yield, S. M. Amirhossein Nejad et al., [18] and S. M. M. Nejad et al., [23] adopted CNN model that face challenges like overfitting, flexibility and the performance is limited when applied to high dimensional data. Addressing these limitations is crucial for developing more robust, accurate, and efficient models capable of supporting sustainable farming practices and decision-making in diverse agricultural contexts.

2.3. An outline of M-Let and CNN –based Fusion model for Crop Yield Prediction

Crop yield prediction is a critical aspect of modern agriculture, aiming to forecast the amount of crop produced in a given season or year. Accurate yield predictions enable policymakers, farmers, and agricultural businesses to make informed decisions regarding resource allocation, market strategies, and food security. With the increasing challenges posed by climate change, soil degradation, and unpredictable weather patterns, the ability to predict crop yields with high precision has become even more crucial. Hence, this paper proposes an innovative M-LNet and CNN –based Fusion model for Crop Yield Prediction (MCF-CYP) model.

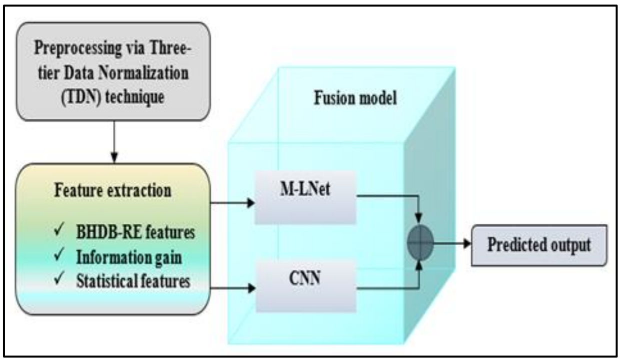


Figure 1 Outline of MCF-CYP model

As depicted in Figure 1, the input data is preprocessed, where a Three-tier Data Normalization (TDN) technique is applied to standardize the data, ensuring all features are on an equivalent scale for better model performance. Following this, feature extraction takes place, where key characteristics of the data are identified. In this case, Beckenstein Hawking with Deng Belief-based Renya Entropy (BHDB-RE) features, information gain, and various statistical features are extracted to capture the most relevant information about the crop yield. These extracted features are then fed into the prediction model, which in this case involves two deep learning architectures: Multi-head LeNet (M-LNET) and Convolutional Neural Networks (CNN). Both models are used to analyze the complex patterns in the data and provide an accurate prediction of the crop yield. Finally, the average of the model outputs the predicted crop yield, which can be used to make informed decisions regarding crop management, and other agricultural practices.

2.4. Preprocessing

The preprocessing step is a crucial part of any data analysis pipeline, especially crop yield prediction. The primary objective is to prepare and clean the raw input data to ensure it is suitable for further processing and modeling.

Assume that the input data is assigned as, which represents the crop yield dataset. This is subjected to preprocessing step, where the data is normalized using Z-score Normalization [28] process. This process is used to transform data so that it has a mean of 0 and a standard deviation of 1. The standard form of Z-score normalization is formulated as in Eq. (1).

$$Zs_i^N = \frac{Zs_i - \mu\left(Zs_i\right)}{\sigma\left(Zs_i\right)} \tag{1}$$

where, indicates matching scores of data; denotes arithmetic mean and refers to standard deviation.

This is particularly useful when the features in a dataset have different units or scales, as it guarantees that each feature contributes similarly to the analysis and avoids biasing the model towards features with larger magnitudes. However, the absolute value of each feature is not considered. This can create ambiguities that have a considerable impact on classifier efficiency. To avoid this issue, a Three-tier Data Normalization (TDN) technique is proposed that eliminates the above shortcomings.

Three-tier Data Normalization: The process of preprocessing the data ${}_{cy}D$ using TDN technique follows the following steps:

Step 1: Assume that the vector representation of an input data as ${}_{cy}D = \left({}_{cy}D_1, {}_{cy}D_2, \ldots, {}_{cy}D_N\right)$ with size N . Initially, the input data be transformed into processed data by multiplying the scaling weight factor with normalized data as defined in Eq. (2).

$$P_D = x(y) \times N({}_{cy}D) \tag{2}$$

where, $\chi(y)$ represents the scaling weight based on input ${}_{cy}D$ such that $\chi(y) > 0$;refers to product operation and $N\left({}_{cy}D\right)$ denotes normalized data that is computed as in Eq. (3).

$$N\left({}_{cy}D\right) = \frac{{}_{cy}D - \mu}{\sigma} \tag{3}$$

Here, $\mu = \frac{1}{N}\sum_{i=1}^{N} {}_{cy}D_i$ and $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left({}_{cy}D_i - \mu\right)^2}$.

Step 2: Subsequently, each column of processed data was normalized based on Improved Z-score Normalization, as given in Eq. (4), which generates new data.

$$P_D' = \left\{\left[\frac{P_D - \mu(P_D)}{\sigma(P_D)}\right] - \left[median\left({}_{cy}D, P_D\right)\right]\right\} \tag{4}$$

Step 3: Each row from the new data was normalized using Double sigmoid normalization [29] as defined in Eq. (5).

$$SS = \begin{cases} \dfrac{1}{1 + \exp\left(-2\left(P_D'/q_1\right)\right)} & if \quad P_D' < t \\ \dfrac{1}{1 + \exp\left(-2\left(P_D'/q_2\right)\right)} & if \quad P_D' \geq t \end{cases} \tag{5}$$

where, denotes double sigmoid normalization; and represents parameter.

Step 4: Compute the average of Double sigmoid normalization of by using Contra harmonic mean as in Eq. (6).

$$SS' = \left[\frac{\frac{1}{m}\sum_{i=1}^{m} SS_i^2}{\frac{1}{m}\sum_{i=1}^{m} SS_i}\right] \tag{6}$$

Further, multiply this average value with the normalized column vectors to obtain normalized data P_D''. Finally, apply tanh normalization [29] to normalized data to obtain three-tire normalized data as in Eq. (7).

$$N_{cy}D = 0.5\left\{\tanh\left(0.001\left(\frac{P_D" - \mu\left(P_D"\right)}{\sigma\left(P_D"\right)}\right)\right) + 1\right\} \quad (7)$$

Thereby, the proposed TND approach consider both the interdependencies between the absolute values of features and the features of each vector of given dataset and this will enhance the efficacy of data mining when executing prediction tasks. Thus, the preprocessed data is represented as.

2.5. Feature extraction

Feature extraction is a critical step in preparing data for ML models, especially in complex tasks like crop yield prediction. The main goal is to identify and extract appropriate features from the preprocessed data that will help the model make accurate predictions.

2.5.1. Beckenstein Hawking with Deng Belief-based Renya Entropy features

Renya defines the information measure in a more generalized form that maintains the additivity of individual events. The Renya entropy [30] is formulated as in Eq. (8). Here, is the Renya entropy at variable and allows uncertainty measurement between the distribution.

$$\mathrm{Re}_b\left(X\right) = \frac{1}{1-b}\log\left(\sum_{i=1}^{n} P_i^b\right) \quad (8)$$

The main limitation of this is often limited in their ability to express complex relationship in high dimensional data. This can be simple measure that

quantify uncertainty but cannot provide rich descriptions of the underlying data structure, which means it cannot capture all the complexities of this relationship making it less effective for prediction.

To address this, a new Beckenstein Hawking with Deng Belief-based Renya Entropy (BHDB-RE) feature extraction process is introduced. At first, the average of Deng Belief entropy [31] measure is computed as in Eq. (9). Further, added to the Beckenstein-Hawking entropy as expressed in Eq. (10). Finally, the Beckenstein Hawking with Deng Belief-based Renya Entropy is formulated as in Eq. (11).

$$DBE\left(g\right) = \frac{\left\{\left[-\sum_{E\subseteq\theta} g\left(E\right)\log_2 \frac{g\left(E\right)+B_m\left(E\right)}{2\left(2^{|E|}-1\right)} . e^{\frac{|E|-1}{|s|}}\right] + \left[\sum_{E\subseteq\theta} g\left(E\right)\log_2\left(2^{|E|}-1\right) - \sum_{E\subseteq\theta} g\left(E\right)\log_2 g\left(E\right)\right]\right\}}{2}$$

$$BH = \frac{k^B c^3}{\hbar G} . \frac{SA}{4} + DBE\left(g\right) \quad (10)$$

$$BHDB^{RE} = \frac{1}{\gamma}\ln\left[1+\gamma . BH\right] \quad (11)$$

where, B_m refers to belief function; g refers to mass function; E denotes focal element of g ; $\left|E\right|$ refers to cardinality of E ; $\hbar$ and k^B represents Boltzmann constant; SA denotes surface area of the sphere ($SA = 4\pi r^2$); r^2 refers to radius; $\hbar = \frac{h}{2\pi}$; h denotes Planck constant; $\left|s\right|$ refers to cardinality of body of evidence(BOE) s , which defines the single element that make up the focal element in BOE s and $\gamma = \left(1-b\right)$.

Thus, the proposed BHDB-RE-based feature extraction captures the complex relationships in the data and the parameter adopted in this, allows tuning the entropy measures based on particular characteristics of data. This approach enables a wider exploration of statistical behaviors.

2.6. Information gain

The information gain (IG) [32] measures the reduction in uncertainty or entropy about a target variable, provided by a feature or attribute. Essentially, it quantifies how much knowing a specific feature helps to reduce the unpredictability of the outcome. The IG is expressed as in Eq. (12).

$$Gn\left(I_j\right) = H\left(P_j\right) - H\left(S_j\right) \tag{12}$$

where, $H\left(P\right) = \sum_{i=1}^{z} P_i \log_2 P_i$. P_i refers to ratio

of conditional parameters P ; The information value $H\left(S_j\right)$ is determined as in Eq. (13), when attribute P_i splits set P using attribute S_j , which has $\left|S_j\right|$ possible attribute values.

$$H\left(S_j\right) = \sum_{i=1}^{z} In_j * H\left(Y_j\right) \tag{13}$$

2.6.1. Statistical features

Statistical features Sf [33] refer to quantitative characteristics derived from preprocessed data $N_{cy}D$ using statistical methods. In this context, the features such as mean, median, standard deviation, minimum and maximum.

Mean: It represents the central tendency of the data, calculated by summing all the values and dividing by the number of observations. It is defined as in Eq. (14).

$$Mn = \frac{1}{v}\sum_{i=1}^{v} z_i \tag{14}$$

Median: The median is the middle value of a dataset when the values are sorted in ascending or descending order.

Standard deviation: The square root of the variance, providing a more interpretable measure of spread in the same units as the data itself. It indicates how much the data deviates from the mean. It is defined as in Eq. (15).

$$Sd = \sqrt{\frac{1}{v}\sum_{i=1}^{v}\left(z_i - Mn\right)^2} \tag{15}$$

- Minimum: The minimum value is the smallest value in the dataset. It provides a lower bound of the data.
- Maximum: The maximum value is the largest value in the dataset.

Thereby, the entire features extracted are symbolized as $Fe = \left[BHDB^{RE}, Gn\left(I_j\right), Sf\right]$.

2.7. M-LNET and CNN-based Fusion model for prediction

The prediction step in machine learning refers to the phase in which a trained model uses its learned patterns to make predictions on new, unseen data. In this work, the models such as Multi-head LeNet (M-LNet) and Convolutional Neural Networks (CNN) are used, in which the both models learn the extracted feature set individually and offers the scores as output. Finally, the average of these scores provides the predicted output, as shown in Figure 2.

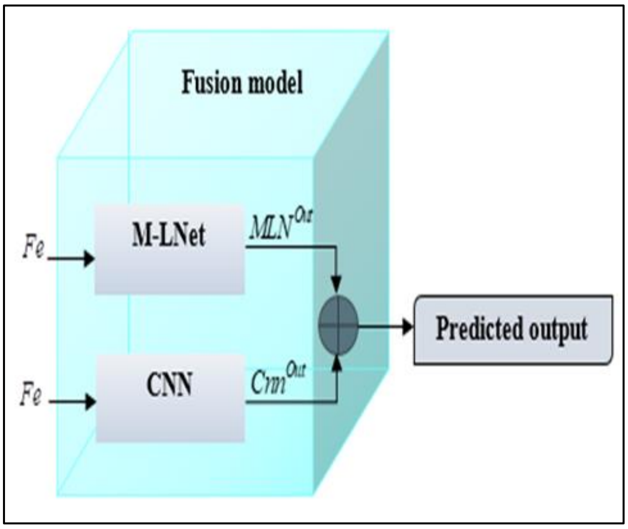


Figure 2 Prediction flow via M-LNET

2.7.1. M-LNET

The Multi-head LeNet is the variant of LeNet architecture [34] that convolves the features into a convolutional layer with six filters, each of size 5 × 5, operating with a stride of 1, no padding, and using the Sigmoid activation function. A convolutional layer of size 5 × 5 with a stride of 1 and no padding results in a receptive field of 5 × 5. However, this is relatively shallow framework with only a few layers. Its performance is constrained when applied to numerous complex and high dimensional data. This will lead to limited capacity, overfitting and flexibility. To avoid these issues, M-LNET model is proposed.

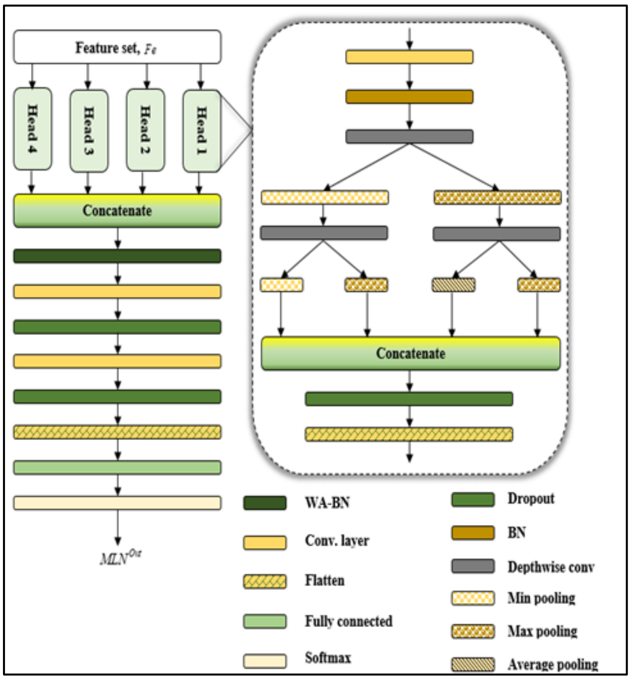


Figure 3 Architecture of M-LNET model

As depicted in Figure 3, the proposed M-LNet model takes input feature set   as input, which is processed through a multi-head LeNet architecture, where the data is passed through four parallel processing heads. Each head begins with an initial convolutional layer (Conv Layer 1) to extract basic spatial features, followed by batch normalization (BN) to stabilize learning and improve training efficiency. After normalization, a depthwise convolution is applied, which helps in capturing channel-specific spatial patterns while reducing computational complexity.

Each head then branches into two different processing paths:

- First Path: It applies min pooling, followed by another depth wise convolution, and then both min and max pooling operations.
- Second Path: It uses max pooling, followed by depth wise convolution, and then both average and max pooling operations.

The outputs of these two paths are then concatenated to combine the extracted features, followed by a dropout layer to prevent overfitting. The resulting feature map is then flattened into a one-dimensional vector.

Once all four heads complete their processing, their outputs are concatenated to integrate diverse feature representations. This combined feature map Weighted Average pooling response Batch Normalization (WA-BN) to maintain stability before passing through an additional convolutional layer for further feature extraction. The WA-BN is the process of applying weight to the average pooling layer [35] that is formulated as in Eq. (16).

$$WA^{Pool} = \left[(1-\delta)\frac{1}{\left|pR_j\right|}\sum_{i\in pR_j} u_i \right] * w_3 \qquad (16)$$

where, w is the weight; $w_3 = 1 - w_1 - w_2$; $w_1 = \cos\theta$; $w_2 = \frac{1}{2}\sin\theta\cos\phi$; $\theta = \frac{2}{\pi}\arccos\frac{1}{3}.\arctan\left(yt\right)$; and $\phi = \frac{1}{2}\arctan\left(yt\right)$.

Further, apply the BN [36] to the weighted average pooling layer as in Eq. (17). Here, $\mu_B\left(x_i\right)$ refers to mean of x_i ; σ_B^2 denotes standard deviation and ε denotes constant.

$$WABn = \left\{ \left[\frac{x_i * WA^{Pool} - \mu_B\left(x_i\right)}{\sqrt{\sigma_B^2 * WA^{Pool} + \varepsilon}} \right] * WA^{Pool} \right\} \, (17)$$

Afterward, a dropout layer is applied to regularize the model, followed by flattening to prepare the data for classification.

Finally, the processed feature vector is passed through a series of fully connected layers, followed by a softmax activation function, which output the final class probabilities as. This architecture efficiently captures multi-scale features and enhances robustness by leveraging multiple parallel processing paths while reducing overfitting through normalization and dropout mechanisms. It improves the stability of training and faster convergence. This helps in diminishing hyperparameter sensitivity, leading to easier and more effective training.

2.7.2. CNN

Convolutional Neural Networks (CNNs) [37] are a type of deep learning model primarily used for tasks such as image recognition, but they can also be applied to other types of data where spatial or hierarchical relationships exist. The architecture of a CNN typically begins with an input layer that takes extracted feature set as input. The input passes through a series of convolutional layers, where several filters are applied to the input feature to detect patterns. Each filter slides over the input, producing a feature map that highlights areas of interest. An activation function is applied after convolution to introduce non-linearity, allowing the model to recognize more sophisticated patterns. The data then passes through pooling layers, usually with max pooling, to down sample the feature map, improving computational efficiency and reducing overfitting. Once the features are retrieved, the data is passed to fully connected (dense) layers, where each neuron is connected to every neuron in the previous layer. These layers help make the final decision, with the last fully connected layer typically using a SoftMax activation function for prediction and provides output as.

3. Results and Discussion

3.1. Simulation Procedure

The proposed Crop Yield Prediction was simulated using PYTHON, specifically "Version 3.7". The processor employed was "11th Gen Intel(R) Core (TM) i3-1115G4 @ 3.00GHz 3.00 GHz and the Installed RAM size was 8.00 GB". Further, the crop yield prediction was analyzed using Crop-Yield-Prediction-in-India dataset [38].

3.2. Dataset Description

The model focuses on predicting the crop yield in advance by analyzing factors like district (assuming same weather and soil parameters in a particular district), "state, season, crop type" using various supervised machine learning techniques. This dataset encompasses a total of 246092,8 data.

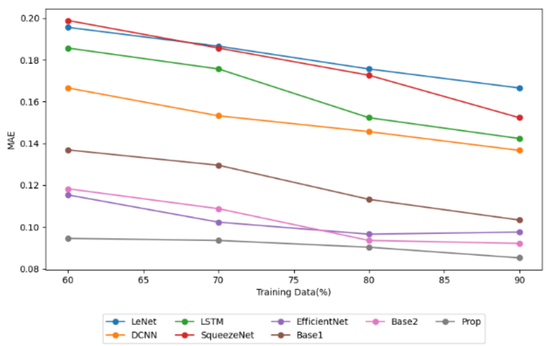
3.3. Performance Analysis

A wide-ranging examination was carried out to assess the efficiency of proposed Crop Yield Prediction in comparison to established approaches. This thorough investigation comprised investigating key measures such as Error Analysis with fluctuating Learning Rates and Training Data, and employed error measures like MAPE, MAE, MSLE, MSE, and RMSE. Furthermore, the evaluation integrated a regression analysis, K-fold cross validation, ablation analysis and statistical evaluations. The Malecon method was compared against state-of-the-art strategies like WTDCNN [26] and BP2, as well as conventional models including LeNet, DCNN, LSTM, Squeeze Net and Efficient Net. The Crop-Yield-Prediction-in-India dataset was utilized for both the simulation and investigation of the Malecon and traditional approaches, providing a robust foundation for the comparative study.

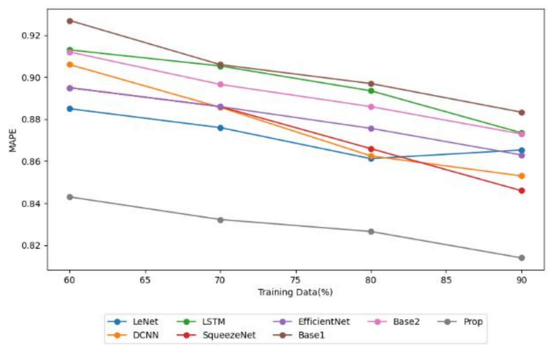
3.4. Error Analysis

The error analysis of Malecon approach for crop yield prediction is represented in figure 4 and it is contrasted with different existing models, such as LeNet, DCNN, LSTM, Squeeze Net, Efficient Net, WTDCNN [26] and BP2. This assessment assesses the performance in terms of MAPE, MSLE, MAE, MSE, and RMSE. Attaining diminished error values across these measures is vital for accurate and consistent prediction of crop yields. At 60% training data, the Malecon approach achieved an MAPE of 0.8483, which enhances to 0.8195 with 90% training data. In contrast, conventional methods like LeNet, DCNN, LSTM, Squeeze Net, Efficient Net, WTDCNN [26] and BP2 demonstrated higher MAPE values, ranging from 0.9289 to 0.8645. In addition, the MSLE achieved using the Malecon scheme is 0.0106 at the training data 90%, while the conventional methods exhibited higher MSLE values with LeNet at 0.0346, DCNN at 0.0257, LSTM at 0.0316, Squeeze Net at 0.0368, Efficient Net at 0.0261, WTDCNN [26] at 0.0285 and BP2 at 0.0235, respectively. The TDN consider both the interdependencies between the absolute values of features and the features of each vector of given dataset. One significant advantage is that the normalized data preserves the interrelationships among an observation's attributes.

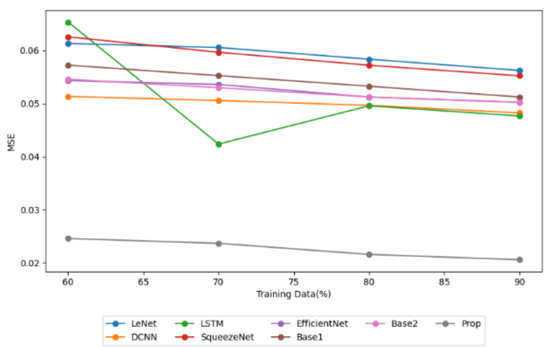
In 70% training data, the Malecon scheme attains an MSE of 0.0251, this is notably lower compared to existing methods. In particular, the MSE for the Malecon approach is substantially better than that of LeNet (0.0624) and DCNN (0.0517). Other methods like LSTM, Squeeze Net, Efficient Net, WTDCNN [26] and BP2 generated the higher MSE scores of 0.0426, 0.0615, 0.0542, 0.0584 and 0.0538, respectively. The lower MSE achieved by the Malecon scheme reflects its superior capability to minimize error values and signifies its efficiency in crop yield prediction. The improved Renya entropy can handle non broad system. The BHDB-RE measure captures the complicated relationships in the data even if high dimensional data. A parameter in our model allows for customizing the entropy measure to suit the distinct features of the data.



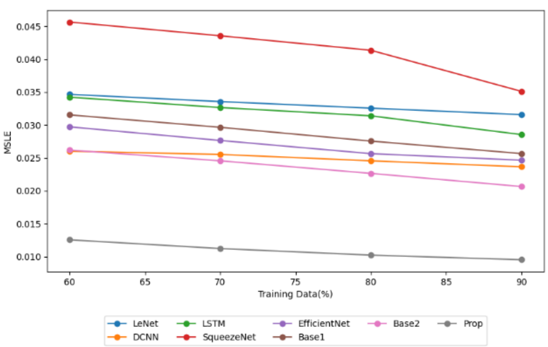
(a)



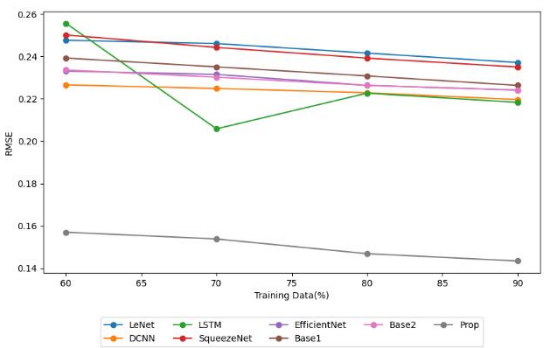
(b)



(c)



(d)



(e)

Figure 4 Error Analysis on M-LNEt+ CNN and Existing Methods a) MAE b) MAPE c) MSE d) MSLE and e) RMSE

3.5. Analysis on Training Loss

Figure 5 explains the training loss curve for M-LNet+CNN model in contradiction to existing methods like LeNet, DCNN, LSTM, SqueezeNet, EfficientNet, WTDCNN [26] and BP2 across 100 epochs for crop yield prediction. All models demonstrate a decreasing loss trend as the number of epochs improves, signifying that they are learning effectively. The LeNet approach exhibit a slightly higher training loss compared to other methods. The M-LNet+CNN model appears to have slightly lower training loss compared to existing methods, suggesting superior performance. Among the models, the EfficientNet, BP2 and SqueezeNet follows a similar trend but do not surpass M-LNet+CNN scheme. All the models show a consistent reduction in training loss, yet the M-LNet+CNN approach perform slightly better than the others.

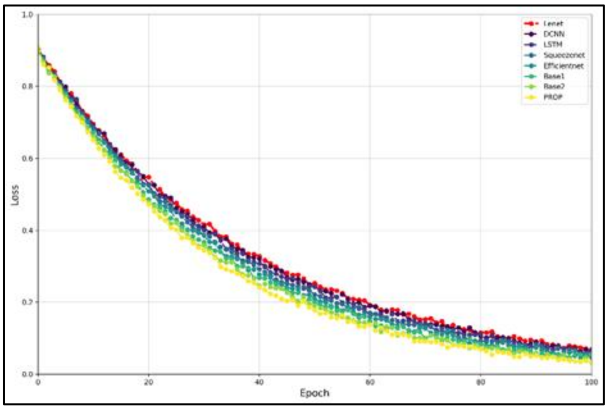


Figure 5 Training Loss Analysis on M-LNet+CNN Vs. Conventional Methods

Statistical Analysis

Table II, Table III, Table IV, Table V and Table VI provide a detailed statistical evaluation comparing the M-LNet+CNN approach against conventional methods for crop yield prediction, such as LeNet, DCNN, LSTM, SqueezeNet, EfficientNet, WTDCNN [26] and BP2. In table VI, the M-LNet+CNN strategy obtained the least RMSE score of 0.1434 in the minimum statistical metric, indicating superior effectiveness in predicting crop yields. In comparison, the EfficientNet and BP2 has the greatest RMSE of 0.2331 and 0.2336, showcasing that its predictions are less accurate than the M-LNet+CNN model. The DCNN, WTDCNN [26] and LeNet have comparatively higher RMSE of 0.2266, 0.2392 and 0.2476. The SqueezeNet and LSTM methods have RMSE scores of 0.2501 and 0.2556, which are notably greater than the RMSE of M-LNet+CNN scheme, suggesting less efficacious prediction abilities. Considering Median Statistical metric, the M-LNet+CNN approach surpasses all established methods with MSLE rate of 0.0107, which is the smallest value recorded among the models. The MSLE values for the conventional strategies are: LeNet (0.0331), DCNN (0.0251), LSTM (0.0320), SqueezeNet (0.0425), EfficientNet (0.0267), WTDCNN [26] (0.0286) and BP2 (0.0236), respectively. The WA-BN in the LeNet model have the ability to enhances the stability of training and faster convergence. It helps to avoid gradient problems. This adjustment lowers reliance on hyperparameter tuning, improving the ease and efficiency of training.

Table 2 Statistical Evaluation on MAE

Statistical Metrics	LeNet	DCNN	LSTM	Squeeze Net	Efficient Net	WTDCNN [26]	BP2	M-LNet+CNN
Mean	0.1811	0.1505	0.1640	0.1774	0.1030	0.1208	0.1032	0.0909
Median	0.1811	0.1494	0.1640	0.1792	0.0999	0.1214	0.1012	0.0919
Standard Deviation	0.0110	0.0110	0.0174	0.0172	0.0075	0.0132	0.0109	0.0036
Minimum	0.1666	0.1367	0.1424	0.1523	0.0966	0.1033	0.0922	0.0852
Maximum	0.1956	0.1666	0.1857	0.1989	0.1153	0.1369	0.1183	0.0945

Table 3 Statistical Evaluation on MAPE

Statistical Metrics	LeNet	DCNN	LSTM	Squeeze Net	Efficient Net	WTDCNN [26]	BP2	M-LNet+CNN
Mean	0.8719	0.8768	0.8964	0.8733	0.8799	0.9033	0.8919	0.8290
Median	0.8707	0.8741	0.8994	0.8760	0.8808	0.9015	0.8913	0.8294
Standard Deviation	0.0093	0.0206	0.0149	0.0189	0.0119	0.0159	0.0143	0.0105
Minimum	0.8612	0.8530	0.8735	0.8460	0.8630	0.8834	0.8730	0.8140
Maximum	0.8850	0.9060	0.9130	0.8950	0.8950	0.9270	0.9120	0.8430

Table 4 Statistical Evaluation on MSE

Statistical Metrics	LeNet	DCNN	LSTM	Squeeze Net	Efficient Net	WTDCNN [26]	BP2	M-LNet+CNN
Mean	0.0591	0.0499	0.0512	0.0587	0.0523	0.0542	0.0523	0.0226
Median	0.0595	0.0501	0.0486	0.0584	0.0524	0.0543	0.0521	0.0226
Standard Deviation	0.0020	0.0012	0.0086	0.0027	0.0017	0.0022	0.0017	0.0016
Minimum	0.0562	0.0482	0.0424	0.0552	0.0502	0.0512	0.0502	0.0206
Maximum	0.0613	0.0513	0.0654	0.0626	0.0543	0.0572	0.0546	0.0246

Table 5 Statistical Evaluation on MSLE

Statistical Metrics	LeNet	DCNN	LSTM	Squeeze Net	Efficient Net	WTDCNN [26]	BP2	M-LNet+CNN
Mean	0.0331	0.0249	0.0317	0.0414	0.0269	0.0286	0.0235	0.0109
Median	0.0331	0.0251	0.0320	0.0425	0.0267	0.0286	0.0236	0.0107
Standard Deviation	0.0011	0.0009	0.0021	0.0039	0.0019	0.0022	0.0021	0.0011
Minimum	0.0316	0.0237	0.0286	0.0351	0.0247	0.0257	0.0207	0.0095
Maximum	0.0347	0.0260	0.0342	0.0457	0.0297	0.0315	0.0262	0.0126

Table 6 Statistical Evaluation on RMSE

Statistical Metrics	LeNet	DCNN	LSTM	Squeeze Net	Efficient Net	WTDCNN [26]	BP2	M-LNet+CNN
Mean	0.2431	0.2235	0.2256	0.2421	0.2288	0.2329	0.2286	0.1503
Median	0.2438	0.2239	0.2205	0.2417	0.2289	0.2329	0.2283	0.1503
Standard Deviation	0.0041	0.0026	0.0184	0.0057	0.0037	0.0048	0.0036	0.0054
Minimum	0.2371	0.2196	0.2058	0.2350	0.2241	0.2263	0.2241	0.1434
Maximum	0.2476	0.2266	0.2556	0.2501	0.2331	0.2392	0.2336	0.1570

3.6. Ablation Study

This study methodically separates elements of a framework to analyze their individual influence on performance. Table VII provides an ablation evaluation on MCF-CYP approach against several alternative variations like MCF-CYP without Feature Extraction, MCF-CYP with Existing Renya Entropy, MCF-CYP with Existing Normalization and MCF-CYP without Normalization for Crop Yield Prediction. For MAE metric, the MCF-CYP approach has the least value of 0.0852, compared to 0.1666 for MCF-CYP without Feature Extraction, 0.1757 for MCF-CYP with Existing Normalization, 0.1296 for MCF-CYP with Existing Renya Entropy and 0.1757 for MCF-CYP without Normalization. This signifies that the MCF-CYP scheme attains the best MAE score in crop yield prediction. When analyzing MSE, the MCF-CYP approach again excels with a notably lower score of 0.0206, while the MCF-CYP without Feature Extraction, MCF-CYP with Existing Normalization, MCF-CYP with Existing Renya Entropy and MCF-CYP without Normalization exhibited greater MSE values. The MCF-CYP with all the improvements like TDN technique, BHDB-RE feature and M-Let makes the MCF-CYP strategy more efficacious in predicting crop yields.

Table 7 Ablation Analysis on MCF-CYP Approach, MCF-CYP with Existing Normalization, MCF-CYP with Existing Renyi Entropy, MCF-CYP without Feature Extraction and MCF-CYP without Normalization

Metrics	MCF-CYP	MCF-CYP without Feature Extraction	MCF-CYP with Existing Normalization	MCF-CYP with Existing Renya entropy	MCF-CYP without Normalization
MSLE	0.0095	0.0260	0.0327	0.0297	0.0326
MAPE	$8.14\times e^{-01}$	$9.06\times e^{-01}$	0.9054	$9.06\times e^{-01}$	0.8612
MAE	0.0852	0.1666	0.1757	0.1296	0.1757
MSE	0.0206	0.0513	0.0424	0.0553	0.0584
RMSE	0.1434	0.2266	0.2058	0.2351	0.2416

3.7. Analysis on K-Fold Cross Validation

K-fold cross-validation involves splitting the data into k equally sized folds. The model is validated k times and trained, with each fold serving as the validation set once while the other k − 1 folds are used for learning. Table VIII explains the K-fold cross validation analysis on Malecon approach is compared with LeNet, DCNN, LSTM, Squeeze Net, Efficient Net, WTDCNN [26] and BP2 for crop yield prediction. The Malecon scheme achieved the least 0.0223 at the 3rd fold, 0.0243 at 4th fold,

MAE score of 0.0216 at the 2nd fold and 0.0233 at the 5th fold. In contrast, the conventional methods like WTDCNN [26], BP2, DCNN, EfficientNet, LeNet, LSTM and SqueezeNet accomplished higher MAE values with greater variations in error values between the folds. The smaller variation in M-LNet+CNN scheme's MAE values across the folds indicates an improved generalization capability, signifying that it is less prone to overfitting.

Table 8 K-Fold Cross Validation Assessment on M-LNet+CNN Versus Extant Methods in terms of MAE

K-Folds	LeNet	DCNN	LSTM	Squeeze Net	Efficient Net	WTDCNN [26]	BP2	M-LNet+CNN
Fold=2	0.0584	0.0497	0.0496	0.0572	0.0512	0.0533	0.0512	0.0216
Fold=3	0.0594	0.0504	0.0512	0.0583	0.0521	0.0563	0.0534	0.0223
Fold=4	0.0601	0.0521	0.0502	0.0604	0.0532	0.0552	0.0544	0.0243
Fold=5	0.0613	0.0512	0.0525	0.0594	0.0543	0.0542	0.0524	0.0233

3.8. Error Analysis on M-LNet+CNN Model in terms of Learning Rate

The error analysis on M-LNet+CNN approach by modifying the learning rate to 0.001, 0.01 and 0.1 for crop yield prediction is displayed in table IX. This analysis suggests that a lower learning rate of 0.001 yields the best performance, indicated by the least error scores across all metrics. As the learning rate improves, the M-LNet+CNN

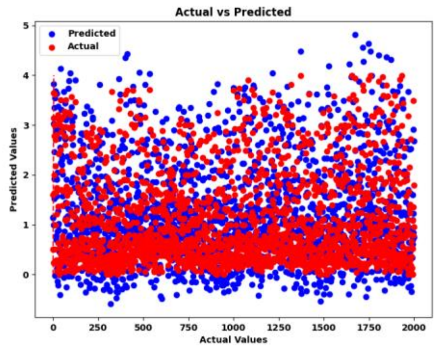
model's predictions become less accurate, resulting in higher errors. For instance, the M-LNet+CNN scheme attained the least RMSE of 0.1434 in the learning rate 0.001, whereas higher learning rates 0.01 and 0.1 resulted in increased RMSE scores of 0.1482 and 0.1502, respectively. Similarly, the least MAPE value is acquired in the learning rate 0.001, this is considerably lower than the other two learning rates.

Table 9 Error Analysis on M-LNet+CNN Strategy by varying the Learning Rate

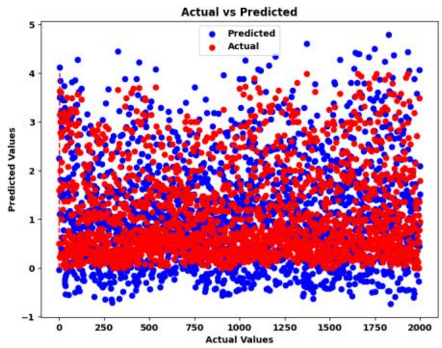
Metrics	Learning Rate=0.001	Learning Rate=0.01	Learning Rate=0.1
MSE	0.0206	0.0220	0.0225
MAPE	0.8140	0.8227	0.8257
MSLE	0.0095	0.0963	0.0976
MAE	0.0852	0.0876	0.0887
RMSE	0.1434	0.1482	0.1502

3.9. Regression Analysis

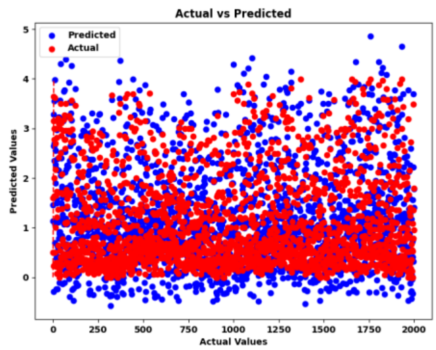
The regression analysis evaluates the performance of M-LNet+CNN approach in comparison to existing methods like WTDCNN [26], BP2, DCNN, EfficientNet, LeNet, LSTM and SqueezeNet for crop yield prediction is shown in figure 6. Effectiveness in this context is determined by how closely a model's predictions match the actual values. Among the models, the M-LNet+CNN method proves superior to conventional approaches, achieving high accuracy with predictions that are either equal to or very near the actual values. In particular, the M-LNet+CNN scheme demonstrates superior prediction ability compared to other methods. Also, it reveals that the M-LNet+CNN approach steadily attains better alignment with actual values compared to the conventional methods. The potential to achieve accurate prediction emphasis that the M-LNet+CNN model's sturdiness and its ability as the most effectual strategy for crop yield prediction.



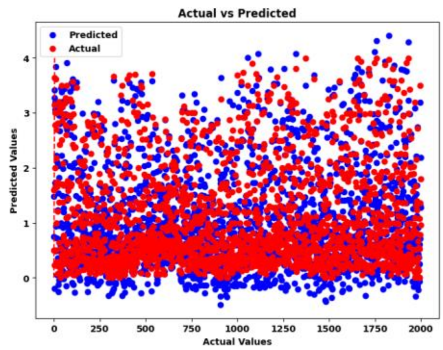
(a)



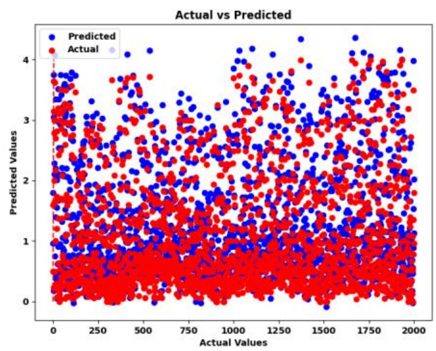
(b)



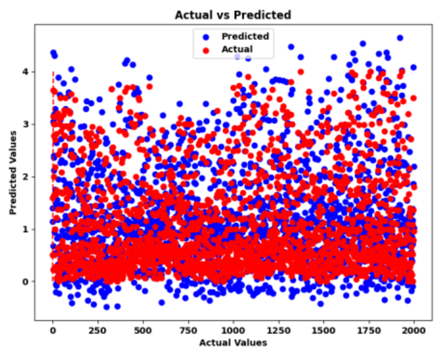
(c)



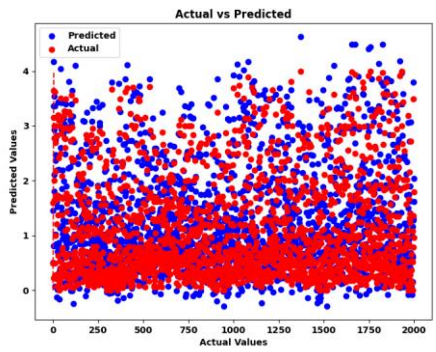
(d)



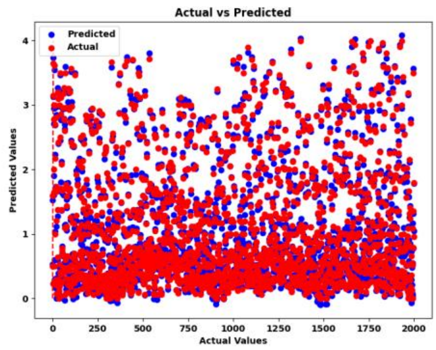
(e)



(f)



(g)



(h)

Figure 6 Regression Analysis a) WTDCNN b) BP2 c) DCNN d) EfficientNet e) LeNet f) LSTM g) SqueezeNet and h) M-LNet+CNN

4. Conclusion

This paper proposed an innovative M-LNet and CNN –based Fusion model for Crop Yield Prediction (MCF-CYP) model. The input data underwent preprocessing, where TDN technique was applied to standardize the values, ensuring all features are on a consistent scale for improved model performance. Next, feature extraction was performed to recognize the key characteristics of the data. In this case, features like BHDB-RE features, information gain, and various statistical measures were extracted to capture essential insights related to crop yield. These features were then passed into the prediction model, which utilizes two deep learning architectures: Multi-head LeNet (M-LNet) and CNN. Both models analyze the complex patterns in the data and generated accurate crop yield predictions. Finally, the predicted yield was outputted, providing valuable insights for crop management and other agricultural decisions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Shabana Ramzan, Yazeed Yasin Ghadi, Hanan Aljuaid, Aqsa Mahmood, and Basharat Ali, "An Ingenious IoT Based Crop Prediction System Using ML and EL", Computers, Materials and Continua, vol. 79, no. 1, 2024.

[2] Muhammad Hanif Tunio, Jian Ping Li, Xiaoyang Zeng, Faijan Akhtar, Syed Attique Shah, Awais Ahmed, Yu Yang and Md Belal Bin Heyat, "Meta-knowledge guided Bayesian optimization framework for robust crop yield estimation", Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 1, 2024.

[3] Dilli Paudel, Allard De Wit, Hendrik Boogaard, Diego Marcos, Sjoukje Osinga, and Ioannis N. Athanasiadis, "Interpretability of deep learning models for crop yield forecasting", Computers and Electronics in Agriculture, vol. 206, 2023.

[4] Malte von Bloh, David Lobell, and Senthold Asseng. "Knowledge informed hybrid machine learning in agricultural yield prediction", Computers and Electronics in Agriculture, vol. 227, 2024.

[5] Giorgio Impollonia, Michele Croci, and Stefano Amaducci, "Upscaling and downscaling approaches for early season rice yield prediction using Sentinel-2 and machine learning for precision nitrogen fertilisation", Computers and Electronics in Agriculture, vol. 227, 2024.

[6] Mahmood Ashraf, Mohammad Abrar, Nauman Qadeer, Abdulrahman A. Alshdadi, Thabit Sabbah, and Muhammad Attique Khan, "A Convolutional Neural Network Model for Wheat Crop Disease Prediction." Computers, Materials and Continua, vol. 75, no. 2, 2023.

[7] Allister Clarke, Darren Yates, Christopher Blanchard, Md Zahidul Islam, Russell Ford, S. Rehman, and Rob Walsh, "The effect of dataset construction and data pre-processing on the eXtreme Gradient Boosting algorithm applied to head rice yield prediction in Australia", Computers and Electronics in Agriculture, vol. 219, 2024.

[8] M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani, "Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion", IEEE Access, vol. 12, pp. 40947-40961, 2024.

[9] Uan Skobalski, Vasit Sagan, Haireti Alifu, Omar Al Akkad, Felipe A. Lopes and Fernando Grignola, "Bridging the gap between crop breeding and GeoAI: Soybean yield prediction from multispectral UAV images with transfer learning", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 210, pp. 260-281, 2024.

[10] Ersin Elbasi, Chamseddine Zaki, Ahmet E. Topcu, Wiem Abdelbaki, Aymen I. Zreikat, Elda Cina, Ahmed Shdefat and Louai Saker, "Crop prediction model using machine learning algorithms", Applied Sciences, vol. 13, no. 16, 2023.

[11] Fatma M.Talaat, "Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes", Neural Computing and Applications, vol. 35, no. 23, pp. 17281-17292, 2023.

[12] Tongxi Hu, Xuesong Zhang, Gil Bohrer, Yanlan Liu, Yuyu Zhou, Jay Martin, Yang Li, and Kaiguang Zhao, "Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield", Agricultural and Forest Meteorology, vol. 336, 2023.

[13] Yuchi Ma, Zhengwei Yang, Qunying Huang, and Zhou Zhang, "Improving the Transferability of Deep Learning Models for Crop Yield Prediction: A Partial Domain Adaptation Approach", Remote Sensing, vol. 15, no. 18, 2023.

[14] Hamzeh F.Assous, Hazem AL-Najjar, Nadia Al-Rousan and Dania AL-Najjar, "Developing a sustainable machine learning model to predict crop yield in the Gulf countries", Sustainability, vol. 15, no. 12, 2023.

[15] Matteo G.Ziliani, Muhammad U. Altaf, Bruno Aragon, Rasmus Houborg, Trenton E. Franz, Yang Lu, Justin Sheffield, Ibrahim Hoteit, and Matthew F. McCabe, "Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model", Agricultural and forest meteorology, vol. 313, 2022.

[16] Hoa Thi Pham,Joseph Awange, Michael Kuhn, Binh Van Nguyen, and Luyen K. Bui, "Enhancing crop yield prediction utilizing machine learning on satellite-based vegetation health indices", Sensors, vol. 22, no. 3, 2022.

[17] S. Vinson Joshua, A. Selwin Mich Priyadharson, Raju Kannadasan, Arfat Ahmad Khan, Worawat Lawanont, Faizan Ahmed Khan, Ateeq Ur Rehman and Muhammad Junaid Ali, "Crop yield prediction using machine learning approaches on a wide spectrum", Computers, Materials and Continua, vol. 72, no. 3, pp. 5663-5679, 2022.

[18] S. M. Mirhoseini Nejad, D. Abbasi-Moghadam and A. Sharifi, "ConvLSTM–ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 17489-17502, 2024.

[19] Uferah Shafi, Rafia Mumtaz, Zahid Anwar, Muhammad Muzyyab Ajmal, Muhammad Ajmal Khan and Zahid Mahmood, "Tackling Food Insecurity Using Remote Sensing and Machine Learning-Based Crop Yield Prediction", IEEE Access, vol. 11, pp. 108640-108657, 2023.

[20] Md Jiabul Hoque, Md. Saiful Islam, Jia Uddin, Md. Abdus Samad, Beatriz Sainz De Abajo and Débora Libertad Ramírez Vargas, "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning", IEEE Access, vol. 12, pp. 47768-47786, 2024.

[21] A. Badshah, B. Yousef Alkazemi, F. Din, K. Z. Zamli and M. Haris, "Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability", IEEE Access, vol. 12, pp. 162799-162813, 2024.

[22] Alakananda Mitra, Sahila Beegum, David Fleisher, Vangimalla R. Reddy, Wenguang Sun and Chittaranjan Ray, "Cotton Yield Prediction: A Machine Learning Approach With Field and Synthetic Data", IEEE Access, vol. 12, pp. 101273-101288, 2024.

[23] S. M. M. Nejad, D. Abbasi-Moghadam, A. Sharifi, N. Farmonov, K. Amankulova and M. Lászlź, "Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 254-266, 2023.

[24] S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers", IEEE Access, vol. 10, pp. 23625-23641, 2022.

[25] Mohammad Amin Razavi, A. Pouyan Nejadhashemi, Babak Majidi, Hoda S. Razavi, Josué Kpodo, Rasu Eeswaran, Ignacio Ciampitti and PV Vara Prasad, "Enhancing crop yield prediction in Senegal using advanced machine learning techniques and synthetic data", Artificial Intelligence in Agriculture, vol. 14, pp. 99-114, 2024.

[26] Leelavathi Kandasamy Subramaniam and Rajasenathipathi Marimuth, "Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops", e-Prime - Advances in Electrical Engineering, Electronics and Energy, vol.8, 2024.

[27] [Che-Hao Chang, Jason Lin, Jia-Wei Chang, Yu-Shun Huang, Ming-Hsin Lai and Yen-Jen Chang, "Hybrid Deep Neural Networks with Multi-Tasking for Rice Yield Prediction Using Remote Sensing Data", Agriculture, vol. 14, 2024.

[28] Aksu, Gökhan, Cem Oktay Güzeller, and Mehmet Taha Eser, "The effect of the normalization method used in different sample sizes on the success of artificial neural network model," International journal of assessment tools in education 6, no. 2, pp.170-192, 2019.

[29] Khalifa, Anouar Ben, Sami Gazzah, and Najoua Essoukri BenAmara, "Adaptive score normalization: a novel approach for multimodal biometric systems," International Journal of Computer and Information Engineering, vol.7, no. 3, pp.376-384, 2013.

[30] Komatsu, Nobuyoshi. "Cosmological model from the holographic equipartition law with a modified Rényi entropy." The European Physical Journal C, vol.77, no. 4, 2017.

[31] Yan, Hangyu, and Yong Deng, "An improved belief entropy in evidence theory," IEEE Access, vol.8, pp.57505-57516, 2020.

[32] Azhagusundari, B., and Antony Selvadoss Thanamani, "Feature selection based on information gain." International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol.2, no. 2, pp.18-21, 2013.

[33] Esmael, Bilal, Arghad Arnaout, Rudolf Fruhwirth, and Gerhard Thonhauser, "A statistical feature-based approach for operations recognition in drilling time series," International Journal of Computer Information Systems and Industrial Management Applications, vol.4, no. 6, pp.100-108, 2012.

[34] Balasubramaniam, Sathiyabhama, Yuvarajan Velmurugan, Dhayanithi Jaganathan, and Seshathiri Dhanasekaran, "A modified LeNet CNN for breast cancer diagnosis in ultrasound images," Diagnostics, vol.13, no. 17, 2023.

[35] Gholamalinezhad, Hossein, and Hossein Khosravi, "Pooling methods in deep neural networks, a review," arXiv preprint arXiv:2009.07485, 2020.

[36] Ioffe, Sergey, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.

[37] Ghosh, Anirudha, Abu Sufian, Farhana Sultana, Amlan Chakrabarti, and Debashis De. "Fundamental concepts of convolutional neural network", Recent trends and advances in artificial intelligence and Internet of Things, pp.519-567, 2020. https://github.com/ankitaS11/Crop-Yield-Prediction-in-India-using-ML/tree/main