EXPLORING INNOVATIVE DESIGN METHODS THROUGH MACHINE LEARNING TECHNIQUES AND COMPARATIVE ANALYSIS WITH EXISTING MODELS

Anshul and Randeep Singh
Department of Computer Science & Engineering,IEC University Baddi, 174103 (Himachal Pradesh)
anshulsheoran91@gmail.com and randeeppoonia@gmail.com

ABSTRACT

The majority of agricultural products have been severely impacted in terms of their performance because of the effects of climate change in India over the last two decades. If policymakers and farmers could have an accurate estimate of the crop yield before it was harvested, it would be easier for them to take the necessary steps to sell and store their produce. Before they begin cultivating their land, the farmers will be able to determine the potential yield of their crop thanks to this research, which will assist them in making the most informed choices possible about their work. It does so by constructing a model of an interactive prediction system to find a solution to the problem. The farmer will have access to the findings of the forecast after it has been completed. Thus, there are a variety of procedures or algorithms that may be used for this sort of data analytics in crop prediction. With the assistance of these algorithms can make predictions about crop production. The findings indicate the mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE) and R^2 have been calculated and the proposed model obtained the RMSE, MSE, MAE, and R^2 values which give the outperformance as compared to other models.

Keywords: Crop yield, Machine learning, ANN, SVM, Random forest.

1. INTRODUCTION

India's economy is heavily dependent on agricultural output because of the country's status as an agrarian society. Hence, India's agricultural sector serves as the country's economic foundation. Agriculture and its related industries, which include forestry and fisheries, contributed 14.5% to India's GDP in 2015 and employed about 50% of the country's workforce [1]. While agriculture's contribution to India's GDP has slowed significantly in recent years, it remains the country's most diverse economic sector and an important part of the country's social and economic framework [2]. Agriculture is the most pervasive industry in terms of population and has a significant impact on India's economy and society as a whole. Production has become considerably more important as time goes on [3]. Changes in precipitation and temperature patterns are a result of the proliferation of new technology and the excessive use of nonrenewable energy sources. As a result of global warming's unpredictability, farmers have a harder time establishing accurate weather forecasts, which in turn reduces agricultural yields. Climate, terrain, history, geography, biology, politics, institutions, and socioeconomics all have a role in shaping India's agricultural landscape [4]. Due to changes in environmental conditions and technological advancements, regulations shifted throughout time. Thus, there are significant variations in agriculture's performance throughout the country's many regions. There is no causal relationship between the several agriculturally relevant elements. Thus, this poses a threat and disrupts the reliable production of food [5].

Machine learning entails imparting information to a machine. Both supervised and unsupervised learning are examples of machine learning methods [6]. In supervised learning, a single human oversees a system while it learns from preexisting data and information in the form of training instances. Supervised learning includes a wide variety of methods such as Bayesian networks, artificial neural networks, support vector machines, decision trees, ID3, hidden Markov models, k-nearest neighbor, and many more. Unsupervised machine learning is providing software with a large quantity of data and allows the algorithm to discover the underlying patterns and relationships [7]. So, employing unsupervised learning to analyze the data might help uncover patterns that were previously concealed. K-NN and partial-based clustering, self-organizing map, hierarchical clustering, k-means clustering, etc. are all instances of unsupervised learning algorithms. Further examples include hierarchical clustering and hierarchical clustering [8].

1.1 Importance of Machine Learning in Crop Yield Prediction

The conventional approach, which is based on data from manual surveys and historical knowledge of the years that came before them, is helpful for a small field of area, but it is difficult to predict for bigger regions or nations using this model. The most recent developments in technology have made data collecting, processing, and storage far more effective than they were in the past. The fast expansion of agriculture necessitates the development of unique processing systems that are capable of managing complicated data. When it comes to data and agricultural research, Machine Learning (ML) has shown its strong performance, particularly when crop categorization and crop yield prediction are taken into consideration [9]. The combination of data analysis and ML creates new opportunities to get a deeper understanding of agriculture and to conduct more thorough field research. ML approaches are more successful for noisy data, and they have the potential to explain non-linear associations. Also, it helps farmers to prepare in advance by employing forecasts [10][11]. The effectiveness of the algorithms improves with the rise in the quantity of input data; as a result, machine learning can handle an enormous amount of data and provide valuable information Figure 1 [12].

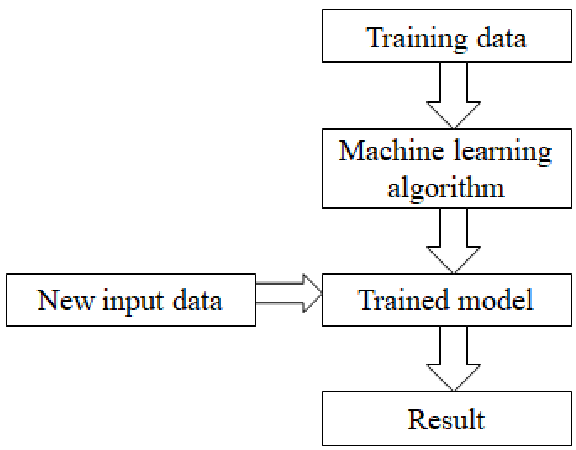

Figure1: Machine learning approach.

In agriculture, crop yield, also known as agricultural output, is an important component to fulfill the requirement of a rising population, and ML is commonly employed to categorize and forecast the yield. The amount of crop yield or productivity that may be achieved in agriculture is contingent on a wide range of parameters, including the state of the soil, the temperature, the amount of precipitation, and the algorithm that is used. Prediction is one use for several of the machine learning methods, while categorization is another [13].

1.2 Factor Affecting the Crop Production

Several different elements have a very significant impact on the production of agricultural products.

- **Variation of Crop Yield with Rainfall**

The amount of precipitation that an area receives significantly affects crop yield. Variations in the amount and pattern of rainfall may have an impact on agricultural productivity. If there is a lot of precipitation, this might result in a decrease in productivity. The amount of rain should not be too severe. The optimal amount of precipitation for crop growth is between 300 and 600mm, which may either result in significant output or ordinary production [14].

- **Variation of Humidity Factor**

The development of agricultural products is substantially influenced by the properties of metrological instruments. Humidity is among the most significant parameters to consider when determining a metrological parameter. Conditions of high humidity are beneficial to the development of any crop.

- **Impact of Climate Change on Agriculture**

The climate is an essential component of agricultural production. Alterations in the climate have a significant impact on the agricultural sector. The amount and timing of precipitation are critical to the success of agricultural endeavors. The patterns of rainfall have shifted as a direct result of changes in the climate. Because of this, the majority of the crops that need water have died. As compared to earlier times, the cost of these crops is much higher. Irrigation has grown more expensive as a result of a lack of rainfall, which has increased the need for tube wells and other such infrastructure [15].

2. REVIEW OF LITERATURE

In this part, various researchers present their results and techniques, expanding on the prior study of a Crop Yield Prediction and Climate Change Impact Assessment Using Machine Learning Technologies in Agriculture.

Gopi et al., (2023)[16] examined that India's agricultural sector is a major contributor to the country's GDP. There are several factors, including soil type and meteorological conditions that must be taken into account when making crop recommendations for a given place. The acreage planted, the kind of irrigation used, the weather forecast, etc. were all factors taken into account while estimating crop production. AI and ML models have recently advanced, paving the door for efficient crop recommendation and crop prediction model development. As such, this research introduces a cutting-edge method for recommending crops and predicting their yields using multimodal machine learning (MMML-CRYP). There are two key functions that the suggested MMML-CRYP model is optimized for: crop recommendation and crop prediction. The first step in effective crop recommendation is the use of an equilibrium optimizer (EO) using the kernel extreme learning machine (KELM) method. The random forest (RF) approach was then used to provide precise predictions about crop productivity. To prove the MMML-CRYP system's efficacy, several simulations were conducted, and the outcomes were compared to a benchmark dataset. Results from experiments showed that MMMLCRYP's highest accuracy was 97.91%, far higher than that of the other techniques tested.

Tamil Sevi et al., (2022)[17] stated that a country's economic growth and development are closely tied to the state of its agricultural sector. Farmers would benefit greatly from accurate predictions of crop output, but doing so is challenging due to the influence of weather and other environmental variables. This cutting-edge approach combines training, classifiers for pre-training, and testing, such as a deep belief network for feature learning, k-means clustering with PSO to provide a global solution and naive Bayes clustering with PSO for evaluation. It is shown that the Lemuria algorithm, when applied to the rainfall dataset, can forecast agricultural yields with an accuracy of 97.74%, and that, when compared to other methods, this yields the best results. This evaluation is carried out in Python.

Mallik et al., (2022)[18] analyzed that the quality of groundwater has suffered due to several factors, both natural and manmade. Thus, it is crucial for the long-term sustainable management of groundwater resources to monitor groundwater quality and determine its appropriateness. The purpose of this investigation was to determine the quality of irrigation water by collecting and analyzing groundwater samples from 35 separate monitoring locations. To find out whether groundwater may be utilized for irrigation, a hybrid MCDM (fuzzy-AHP) approach was applied. Using spatial overlay analysis, a map of areas with varying degrees of appropriateness for irrigating crops was created. To forecast whether or not irrigation water would be suitable, many different regression-based machine-learning models were utilized and evaluated. The results showed that the best model was an ANN with an R2 of 0.990 and an RMSE of almost zero (0). In a place where routine sampling and analysis offer significant difficulties, the current approach may help predict irrigation water suitability.

Elavarasan et al., (2021)[19] intended that there would inevitably be a massive amount of data from many agricultural domains made available to the public as a result of the advancement of research and innovation. New ideas in the many subfields of agriculture may now be conceived with the help of machine learning and its accompanying processing power. It would be interesting to see whether we could develop a method of predicting agricultural yields based on variables such as weather, soil, and water availability. Using a combination of deep belief networks (DBN) and fuzzy neural networks (FNN), this research suggests a method for predicting agricultural yields using deep learning. In DBN, neural networks and statistical analysis are brought together. Nonlinear systems benefit more from DBN's performance, but the algorithm itself cannot deliver on expectations for resilience, model correctness, or learning speed. Thus, it has been suggested to use a DBN in conjunction with an FNN to fix the issues of nonlinearity and gradient diffusion. For better model development and feature vector creation, the suggested model begins with a DBN pre-training procedure. The FNN is given this distinctive feature vector as an input for analysis. By comparing the proposed fuzzy neural network-based deep belief network with competing deep learning algorithms, its superiority is examined. The suggested model outperforms competing models in terms of predictive accuracy (92%), while still maintaining the integrity of the original data distribution.

Khaki et al., (2021)[20] suggested that it is now feasible to estimate agricultural yields on a global scale with the use of remote sensing data, which allows for round-the-clock monitoring of crops. Stakeholders may optimize production potential by making choices in real-time with this data. While several models exist for extrapolating yield from remote sensing data, there is not yet a method that can estimate yield for numerous crops at once, which would allow for more precise extrapolations. We present YieldNet, a novel convolution neural network model that makes use of a deep learning framework that employs transfer learning across forecasts of corn and soybean yields by reusing the weights of the underlying feature extractor. Moreover, we suggest a novel loss function to take into account the multi-target response variable. We collect information from 1132 corn-growing counties and 1076 soybean-growing counties throughout the United States for our project. Numerical findings show that our suggested strategy is competitive to existing state-of-the-art methodologies, with an MAE of 8.74% and 8.70% of the average yield for maize and soybeans, respectively, between one and four months before harvest, respectively.

Mupangwa et al., (2020)[21] examined that research methods like crop simulation models are used extensively in the agricultural sciences to supplement field trials and investigate the effects of new technology. New generations of artificial intelligence (AI) technologies, known as machine learning (ML) methods, provide a viable alternative to and supplement conventional agricultural production models. In the current research, linear algorithms (LDA and LR) were able to provide more accurate predictions of maize production compared to nonlinear techniques (NB, KNN, CART, and SVM). Yield prediction using the KNN algorithm, however, was on par with that using the linear tools evaluated in this work. The LDA method performed the best overall, whereas the SVM algorithm performed the worst for predicting maize yield. Before putting machine learning (ML) techniques into use in agriculture, it is essential to evaluate their effectiveness using many criteria.

Archana et al., (2020)[22] studied that agriculture is the foundation of every nation that is still growing, such as India. The bulk of their people is involved in agriculture since it is their primary source of income. The discipline of machine learning is an emerging subfield of informatics that has the potential to be applied to agriculture in a way that is both effective and efficient. The ability to estimate and forecast crop output is crucial for those involved in agriculture, and it may be obtained via the use of machine learning methods. When farmers aren't aware of the nutrients in the soil or the makeup of the soil, it leads to a low crop output. As a result, the suggested method was established, which concentrates on the macronutrients (NPK), pH, and electrical conductivity in the soil, as well as temperature, to provide the most suitable crop ideas. Crop rotation, crop output prediction and forecasting, and fertilizer recommendation are all elements that would be included in the system that is being suggested here. Throughout this project, a system would be constructed that makes use of the agricultural dataset and applies the voting-based ensemble classifier algorithm to provide recommendations for the most suitable crops. The ability to estimate and forecast crop yields will increase agricultural productivity. The fertility of the soil may be increased by the rotation of crops at regular intervals. The technology helps farmers make decisions about fertilization that are in their best interests. The accuracy of this technique was somewhere in the vicinity of 92%.

Medar et al., (2019)[23] stated that the agricultural sector is crucial to the growth of the national economy. Agricultural practices were the catalyst for the development of human culture. India's economy relies heavily on agricultural output because of the country's status as an agrarian society. Because of this, agriculture has the potential to serve as the foundation of the nation's economy. Crop selection is a crucial part of agricultural planning. Market price, production rate, and government regulations are only some of the variables that would guide crop choices. The Indian economy desperately needs widespread reforms in the agricultural sector. Machine learning methods may be readily used in the agricultural sector, allowing us to make significant strides forward in this important industry. Improvements in agricultural machinery are important, but so is access to relevant and up-to-date knowledge on a variety of topics. The central idea of this research is to use the crop selection approach to improve agricultural practices and reduce farmers' workloads. This boosts the Indian economy by increasing the agricultural output per acre.

Gopal et al., (2019)[24] examined that numerous studies have been conducted in the agricultural arena utilizing the ML algorithm ANN and the statistical model Multiple Linear Regression (MLR) to forecast higher crop output. In this post, we take a look at how MLR and ANN are intrinsically linked. In this study, we present a hybrid MLR-ANN model for accurate agricultural production forecasting. The input layer weights and bias of the ANN were initialized using the MLR intercept and coefficients, and this hybrid model examines the resulting prediction accuracy. A feed-forward ANN was trained using a back-propagation technique to provide precise predictions about rice crop productivity. The MLR's coefficients and bias are used to establish the input layer weights and bias in this MLR-ANN hybrid model, rather than random values. The prediction accuracy of the hybrid model is measured against that of ANN, MLR, SVR, K-NN, and RF models using these performance metrics. Both hybrid MLR-ANN and traditional ANN's computational times were determined. As shown in the findings, the suggested hybrid MLR-ANN model outperforms the status quo models in terms of accuracy.

Sun et al., (2019)[25] stated that there are several important applications for yield forecast data, including yield mapping, agricultural market planning, crop insurance, and harvest management. Predicting agricultural yields via remote sensing is becoming more crucial. Machine learning and specifically the DL approach (using techniques like CNN or LSTM) has enabled significant advancements in this area by using remote sensing data. New research in this field suggests that LSTM may disclose phonological traits and that CNN can investigate additional spatial variables, both of which are crucial to crop output prediction. Few trials, however, have been documented that combine these two models for predicting crop yields. This research proposes a deep CNN-LSTM model for predicting CONUS county-level soybean yields after the growing season and throughout the growing season. To train the model, we used historical soybean yield data as labels in addition to data on meteorological and environmental factors such as MODIS Land Surface Temperature (LST) and MODIS Surface Reflectance (SR). All of the GEE-based training data were aggregated and converted into histogram-based tensors for deep learning. The experimental findings show that the proposed CNN-LSTM model's prediction performance can surpass that of the pure CNN or LSTM model at the season's conclusion and throughout the year. Future yield prediction for fine-scale crops like maize, wheat, and potatoes might benefit greatly from the suggested strategy.

2.1 Comparison of Reviewed Technique

The following study expands on the previous performance evaluation of feature-based opinion using the KPCA technique; several researchers explain their findings as seen in table 1 below.

Table1: Comparison of reviewed technique.

Authors [Ref.]	Technique	Outcome
Gopi et al., (2023)[16]	AI and ML	Results from experiments showed that MMMLCRYP's highest accuracy was 97.91%, far higher than that of the other techniques tested.
Tamil Sevi et al., (2022)[17]	Lemuria Algorithm	The Lemuria algorithm, when applied to the rainfall dataset, can forecast agricultural yields

		with an accuracy of 97.74%, and that, when compared to other methods, this yields the best results.
Mallik et al., (2022)[18]	Hybrid MCDM	The results showed that the best model was an ANN with an R2 of 0.990 and an RMSE of almost zero (0).
Elavarasan et al., (2021)[19]	DBN and FNN	The suggested model outperforms competing models in terms of predictive accuracy (92%), while still maintaining the integrity of the original data distribution.
Khaki et al., (2021)[20]	ML	Numerical findings show that our suggested strategy is competitive to existing state-of-the-art methodologies, with an MAE of 8.74% and 8.70% of the average yield for maize and soybeans, respectively, between one and four months before harvest, respectively.
Mupangwa et al., (2020)[21]	LDA	The LDA method performed the best overall, whereas the SVM algorithm performed the worst for predicting maize yield.
Archana et al., (2020) [22]	Ensemble classifier algorithm	The findings indicate that the accuracy of this technique was somewhere in the vicinity of 92%.
Medar et al., (2019)[23]	ML	The central idea of this research is to use the crop selection approach to improve agricultural practices and reduce farmers' workloads.
Gopal et al., (2019)[24]	Hybrid ANN	As shown in the findings, the suggested hybrid MLR-ANN model outperforms the status quo models in terms of accuracy.
Sun et al., (2019)[25]	CNN-LSTM	The experimental findings show that the proposed CNN-LSTM model's prediction performance can surpass that of the pure CNN or LSTM model at the season's conclusion and throughout the year.

3. PROBLEM FORMULATION

A wide range of climate factors affects the agricultural output. Metrological variables (such as humidity) are monitored and recorded much as their meteorological and pedological counterparts (such as precipitation, regional precipitation, irrigation, and temperature) (PH, organic carbon, phosphorus, fiber, etc.). Moreover, everything is chaos due to the ongoing climate change. Despite modern advances, many Indian farmers continue to use time-tested practices. But, back when the environment was milder, everything went smoothly and on time. But today, almost everything has changed because of global warming and other factors. The agriculture industry in India is struggling due to a lack of consistent, timely rainfall. Humidity is beneficial for plants, but the present levels are too high. As a result of the winter season being thrown off, there has been significant damage to Rabi crops. Rainy winters have been the norm for the last several decades. To address these issues, we need a structure that can help us find previously hidden data, results, patterns, and insights. The farmer will have a better idea of what to plant and how to do so that he may get the best possible harvest. The proposed system utilizes data analytics techniques to analyze data on agricultural output, intending to assist farmers in making informed decisions. Hence, we provide a standardized framework based on a machine-learning model for producing accurate yield forecasts.

4. RESEARCH METHODOLOGY

The method used is the most important part of any inquiry. It serves as the central fulcrum around which the whole investigation is centered. To perform an empirical examination of the statistical relationship between climate change and Indian agriculture, the current research effort has been focused on the analysis of secondary data. The term "secondary data" refers to any kind of data that has been gathered in the past and is now available for use by further researchers. This kind of data was selected rather than primary data because the costs involved with gaining access to it are much lower than the costs associated with collecting the information on an individual basis. Assessing prediction algorithms using fewer important features is the goal of the work being done on this subject at the moment. For us to achieve the objectives of the research, we need to carry out the essential steps, the most important of which are the gathering of data and maintaining purity. As a result of the fact that this investigation is a combination of the three previously mentioned methods, the technique that is described in each strategy as well as the various resources have been combined into a single cohesive system to accomplish the goal. The data that were used for this study include data about irrigation, data about meteorology, data about the use of fertilizer, and yield statistics. The data were obtained from a variety of sources and then subjected to pre-processing. The pre-processed data were then run through the algorithms for selecting the most relevant features to establish which characteristics were the most important. The chosen properties of the input dataset are sent to predictive algorithms to use as input to make predictions about crop yield.

4.1 Technique

- **Random Forest**

Artificial intelligence machine learning algorithms may benefit greatly from this particular ensemble technique since it is used to increase both their efficiency and precision. When data from many trees are combined, the resulting predictions have higher accuracy. In Figure 2 we see an illustration of a random forest. A decision tree can only lead to one conclusion and give a limited variety of groups, whereas a decision tree forest can cover any possibilities [26].

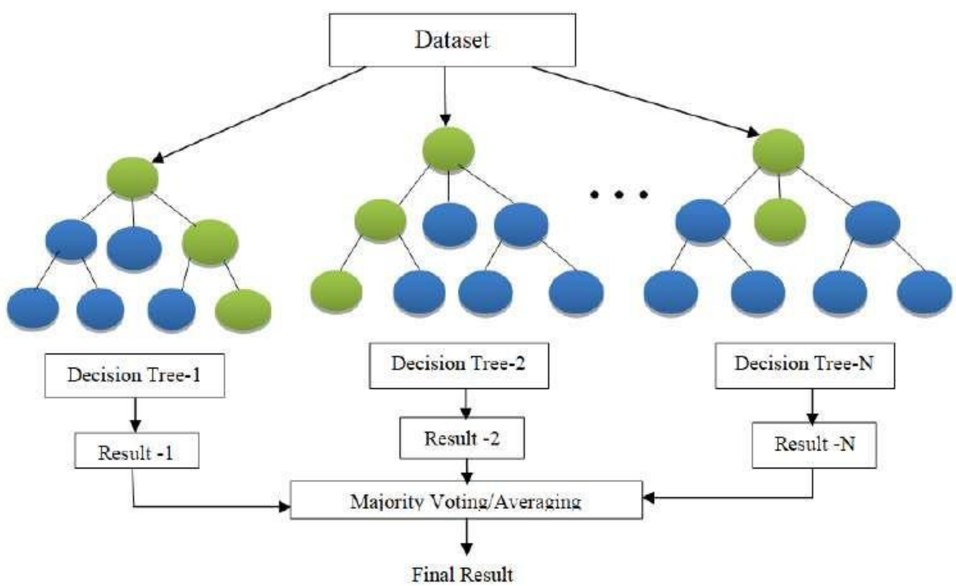

Figure2: Structure of Random forest [27].

- **Information Fuzzy network**

Information fudging networks are a greedy machine technique for supervised learning. The Info Fuzzy Network is a data structure that emerges from the learning process. IFN is built similarly to how decision trees are built. Regardless, IFN generates a graph instead of a circle.

- **Artificial Neural Networks (ANN)**

Data processing models based on the way the brain and other biological nervous systems handle information are called artificial neural networks (ANNs). They are concentrating on the neuronal structure of the cerebral cortex of mammalian organisms, but on a much more minute scale. A significant number of professionals in the field of artificial intelligence are of the opinion that ANN provides the greatest and maybe only opportunity for the development of intelligent machines. Figure 3 depicts the architecture of ANN as shown below [28].

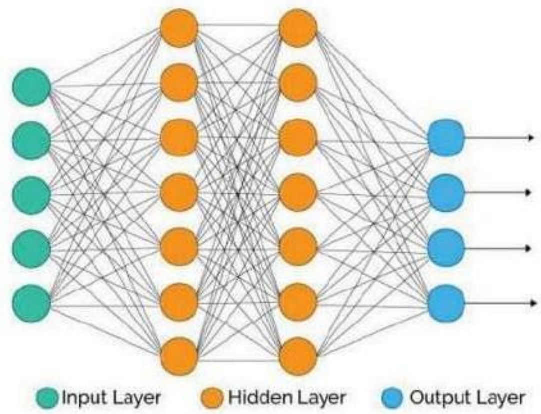

Figure3: ANN architecture [29].

- **Support Vector Machines (SVM)**

Parametric, linear data may be analyzed using time series analysis to glean relevant statistics and other data properties. To foretell future values, analysts use forecast time series. Frequency field analysis and time field analysis are two examples of time series analysis techniques that may be used for both linear and non-linear, univariate and multivariate data. Accurate crop projections cannot be made without first analyzing time series data. Yield may be shown as a function of time to illustrate the relationship between the two variables. The basic structure of an SVM is seen in Figure 4 [30]. The signal vector that is being input is what makes up the input layer. The input signal vector (x) and the support vector are used to calculate an inner-product kernel in the hidden layer (si). A buried layer neuron's linear outputs are added together in the output neuron. There is a bias in the neuron's output.

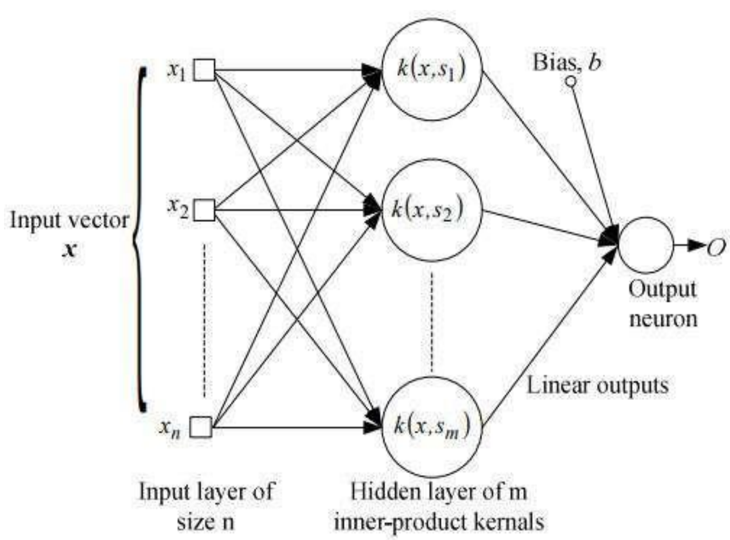

Figure4: The general architecture of SVM [30].

The interim output, O of a support vector machine can be computed as:

$$O = \sum_i w_i k(x_i, s_i) + b \qquad (1)$$

Where si is the support vector, x is the input vector, wi is the weight vector, and b is the bias. The function k(x, si) is a kernel of x and si.

5. PROPOSED FRAMEWORK

The structure of the proposed methodology is shown in figure 5. The following described proposed methodology which as described given below.

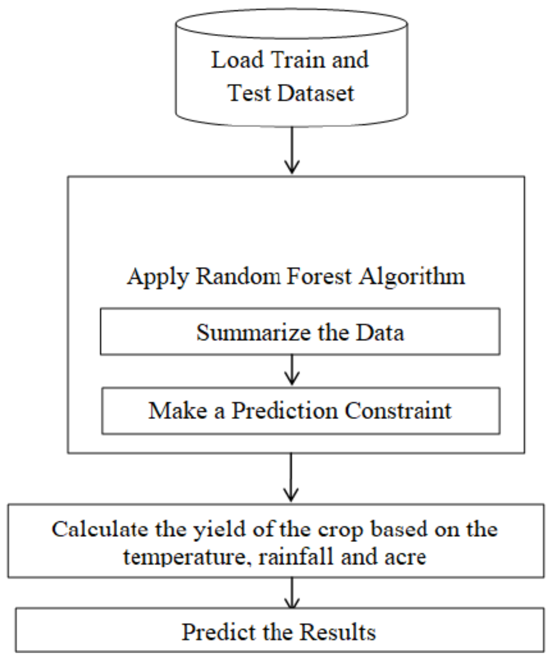

Figure5: Proposed framework.

There are the following steps described in the above mention process.

Step 1: At the beginning, we collect the dataset by gathering the data from a variety of different sources.

Step 2: The data should then be divided into load training and testing datasets; the training dataset should have the highest information density since this will be used to teach the majority of the example datasets how to generate the desired output. The samples for the training set are now being gathered, and the testing set will make use of the remaining information to evaluate the performance of the system.

Step 3: The next step is to apply the random forest method, which will summaries the data and provide a forecast based on the constraints.

Step 4: During this stage of the process, the algorithm will calculate the yield of the crop based on the temperature, rainfall, and acreage, and it will make predictions about the results. Once this stage is complete, the algorithm will not only assist farmers in selecting the appropriate crop to cultivate during the following growing season, but it will also bridge the gap between the technology sector and the agricultural industry.

6. RESULT AND DISCUSSION

Multiple results illustrate the discrepancy between the proposed model and previous research models. In addition, the findings are demonstrating the efficacy of the suggested model in terms of recall, precision, F1 score, and confusion matrix.

Result 1

Figure 6 depicts the confusion matrix for multi-class classification. This matrix is described in terms of both its true label and its predicted label. Count values are used to provide a concise summary of the numbers of accurate and inaccurate predictions.

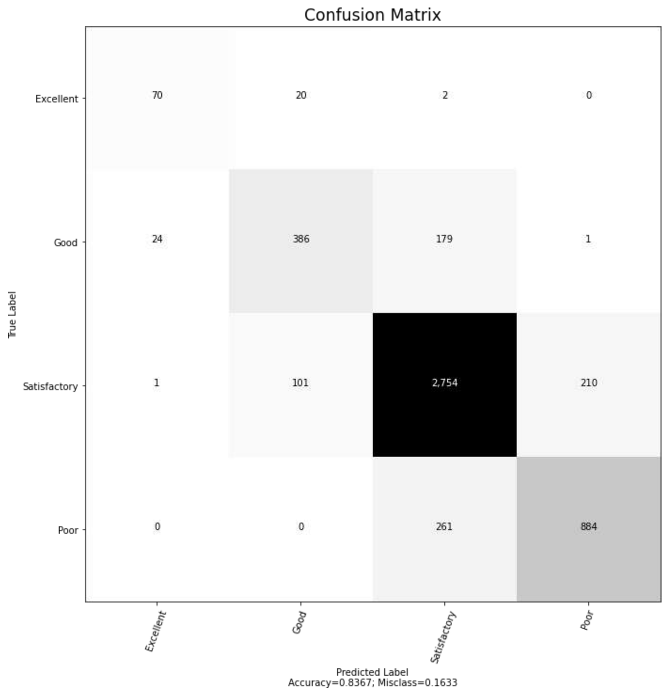

Figure6: Confusion matrix.

Result 2

Figure 7 illustrates the model accuracy and the loss that is shown below which is defined in terms of training and validation. The accuracy of the training is represented by the green line, while the accuracy of the validation is shown by the red line. Both of these accuracies are also reflected in the model loss, which is shown below. From Fig.7, the model accuracy significantly increases while model loss is significantly going down. Model accuracy was a representation of the model's performance, which was defined as a continuous improvement based on epochs in a manner that was easy to understand and computed as a percentage. The loss is stated in terms of between loss and epochs, and its interpretation is defined as being that models are continuously going down in decreasing form.

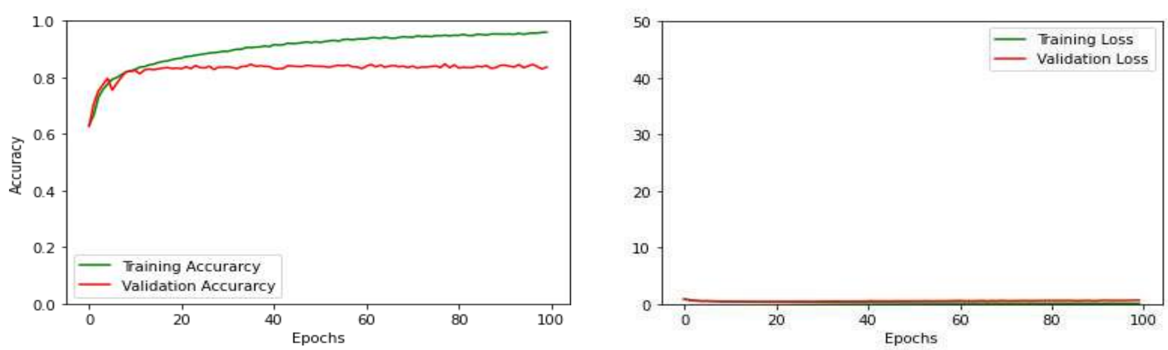

Figure7: Model accuracy and Loss.

Result 3

Figure 8 depict the comparison graph of mean absolute error which is described in the different model as shown below. There are several models which has been consider such as Modified CNN-Bidirectional LSTM model / Proposed model, CNN-DNN (with RobustScaler and SelectFromModel), CNN-XGBoost (with RobustScaler), XGBoost (raw), XGBoost (with RobustScaler), XGBoost (with RobustScaler and SelectFromModel), CNN-LSTM (with RobustScaler and SelectFromModel), and CNN-RNN (with RobustScaler and SelectFromModel). From the figure, it is clear that the XG Boost (raw) has high mean absolute error while proposed method obtained low mean sqaured error as compared to other methods.

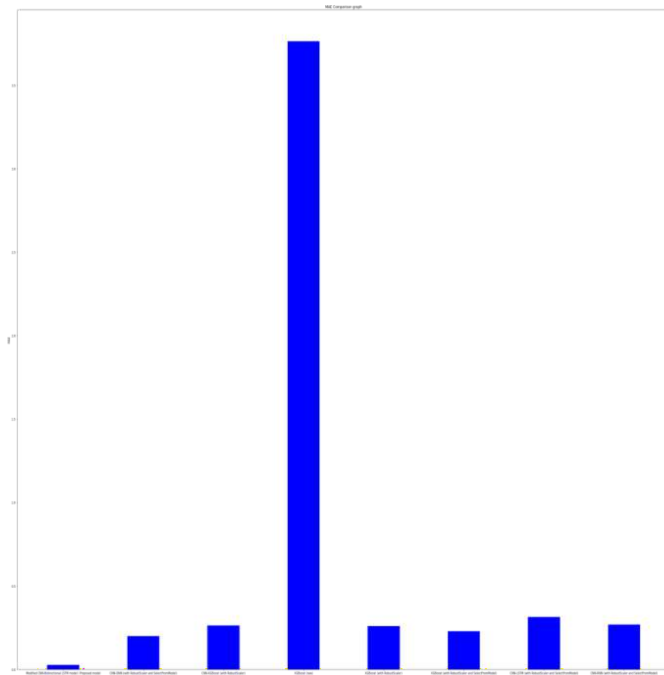

Figure8: Comparison graph of Mean absolute error.

Result 4

Figure 9 depict the comparison graph of mean squared error which is described in the different model as shown below. There are several models which has been consider such as Modified CNN-Bidirectional LSTM model / Proposed model, CNN-DNN (with RobustScaler and SelectFromModel), CNN-XGBoost (with RobustScaler), XGBoost (raw), XGBoost (with RobustScaler), XGBoost (with RobustScaler and SelectFromModel), CNN-LSTM (with RobustScaler and SelectFromModel), and CNN-RNN (with RobustScaler and SelectFromModel). From the figure, it is clear that the XG Boost (raw) has high mean squared error while proposed method obtained low mean sqaured error as compared to other methods.

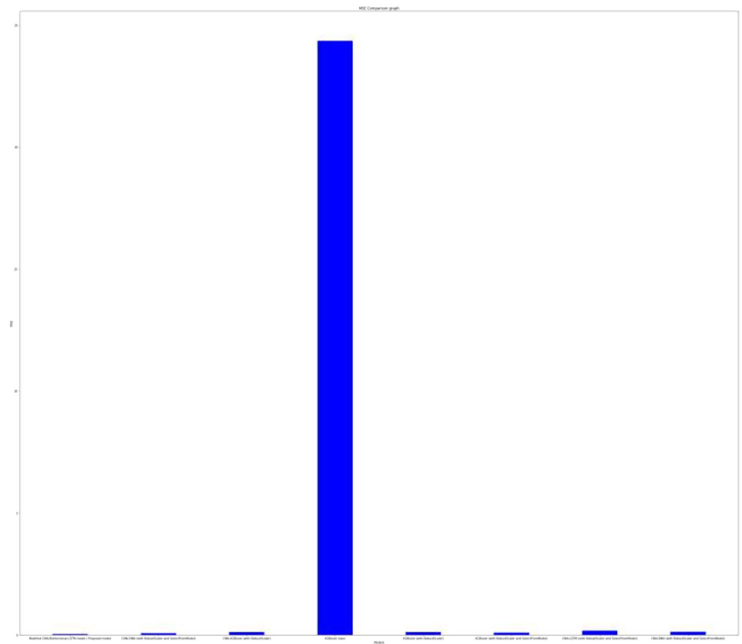

Figure9: Comparison graph of Mean squared error.

Result 5

Figure 10 depict the comparison graph of root mean square error which is described in the different model as shown below. There are several models which has been consider such as Modified CNN-Bidirectional LSTM model / Proposed model, CNN-DNN (with RobustScaler and SelectFromModel), CNN-XGBoost (with RobustScaler), XGBoost (raw), XGBoost (with RobustScaler), XGBoost (with RobustScaler and SelectFromModel), CNN-LSTM (with RobustScaler and SelectFromModel), and CNN-RNN (with RobustScaler and SelectFromModel). From the figure, it is clear that the XG Boost (raw) has high root mean square error while proposed method obtained low root mean square error as compared to other methods.

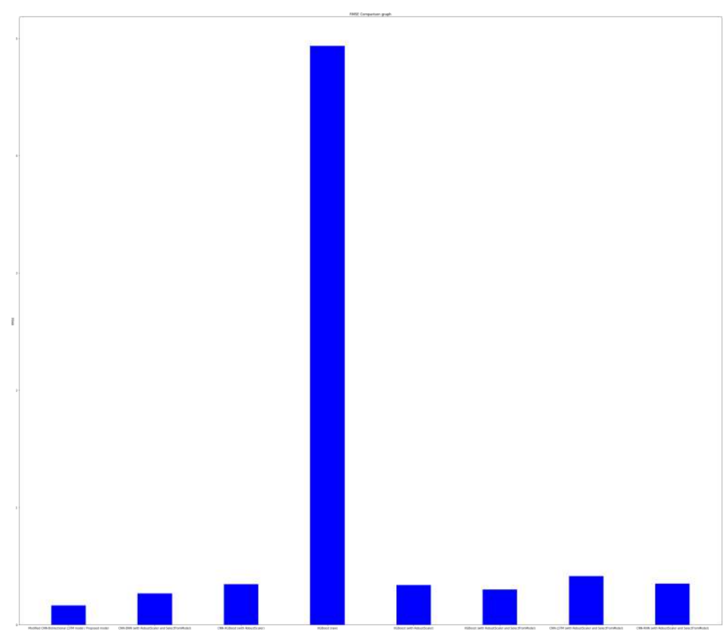

Figure10: Comparison graph of RMSE.

Result 6

Figure 10 depict the comparison graph of R^2 error which is described in the different model as shown below. There are several models which has been consider such as Modified CNN-Bidirectional LSTM model / Proposed model, CNN-DNN (with RobustScaler and SelectFromModel), CNN-XGBoost (with RobustScaler), XGBoost (raw), XGBoost (with RobustScaler), XGBoost (with RobustScaler and SelectFromModel), CNN-LSTM (with RobustScaler and SelectFromModel), and CNN-RNN (with RobustScaler and SelectFromModel). From the figure, it is clear that the CNN-DNN (with RobustScaler and SelectFromModel) has high R^2 error while proposed method obtained low R^2 error as compared to other methods.

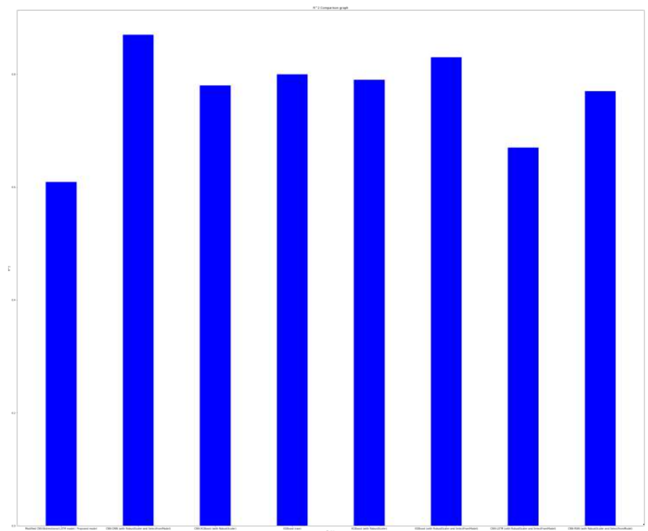

Figure11: Comparison graph of R^2.

Result 7

Figure 12 demonstrates the classification report means described the performance parameters such as precision, recall, f1-score, and support as shown below. The excellent precision is 0.74, recall is 0.76, f1-score is 0.75 and support is 92 as shown below.

	precision	recall	f1-score	support
Excellent	0.74	0.76	0.75	92
Good	0.76	0.65	0.70	590
Satisfactory	0.86	0.90	0.88	3066
Poor	0.81	0.77	0.79	1145
accuracy			0.84	4893
macro avg	0.79	0.77	0.78	4893
weighted avg	0.83	0.84	0.83	4893

Figure12: Performance parameters.

Result 8

Tables 2 illustrate the performance of proposed models concerning different evaluation as shown below. There are seven models is calculated with their values such as RMSE, MSE, MAE, and R^2. From table 2, CNN-DNN (with RobustScaler and SelectFromModel) is calculated and their RMSE value is 0.266, MSE value is 0.071, MAE value is 0.199 and R^2 value is 0.87, the RMSE, MSE, MAE, and R^2 value of CNN-XGBoost (with RobustScaler) is 0.346, 0.120, 0.263, and 0.78 and the RMSE value of the proposed model is 0.16, MSE value is 0.027, MAE value is 0.027 and R^2 value is 0.61 as shown below.

Table2: Performance of proposed model concerning different evaluations.

Models	RMSE	MSE	MAE	R^2
Modified CNN-Bidirectional LSTM model / Proposed model	0.16	0.027	0.027	0.61
CNN-DNN (with RobustScaler and SelectFromModel)	0.266	0.071	0.199	0.87
CNN-XGBoost (with RobustScaler)	0.346	0.120	0.263	0.78
XGBoost (raw)	4.936	24.360	3.764	0.80
XGBoost (with RobustScaler)	0.337	0.113	0.259	0.79
XGBoost (with RobustScaler and SelectFromModel)	0.299	0.089	0.228	0.83
CNN-LSTM (with RobustScaler and SelectFromModel)	0.415	0.172	0.314	0.67
CNN-RNN (with RobustScaler and SelectFromModel)	0.350	0.123	0.269	0.77

7. CONCLUSION AND FUTURE WORK

This study examined crop yield prediction using ML algorithms and explored a range of aspects that rely on data availability. This study presents crop yield and forecast models. ML and the agricultural domain field improved crop prediction in the experimental investigation. From the result, the comparison graph of MSE, RMSE, MAE, and R^2 has been estimated and precision, recall, f1-score, and support parameters have been calculated as mentioned in the above results section. The research described in this study has the potential to be expanded further. The development of machine learning models using climate change and other methodologies is, as far as we are aware, a pioneering effort aimed at boosting agricultural output. The ability of the algorithm to estimate crop production is discussed in this research. In the future, we will be able to pick an accurate algorithm for crop production prediction by determining which algorithm is the most efficient based on the accuracy metrics that are associated with each method.

REFERENCES

[1] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." In *2020 third international conference on smart systems and inventive technology (ICSSIT)*, pp. 926-932. IEEE, 2020.

[2] Bali, Nishu, and Anshu Singla. "Emerging trends in machine learning to predict crop yield and study its influential factors: A survey." *Archives of computational methods in engineering* (2022): 1-18.

[3] Chlingaryan, Anna, Salah Sukkarieh, and Brett Whelan. "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review." *Computers and electronics in agriculture* 151 (2018): 61-69.

[4] Cedric, Lontsi Saadio, Wilfried Yves Hamilton Adoni, Rubby Aworka, Jérémie Thouakesseh Zoueu, Franck Kalala Mutombo, Moez Krichen, and Charles Lebon Mberi Kimpolo. "Crops yield prediction based on machine learning models: the case of west african countries." *Smart Agricultural Technology* (2022): 100049.

[5] Veenadhari, S., Bharat Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." In *2014 International Conference on Computer Communication and Informatics*, pp. 1-5. IEEE, 2014.

[6] Mishra, Subhadra, Debahuti Mishra, and Gour Hari Santra. "Applications of machine learning techniques in agricultural crop production: a review paper." *Indian J. Sci. Technol* 9, no. 38 (2016): 1-14.

[7] Cunha, Renato LF, Bruno Silva, and Marco AS Netto. "A scalable machine learning system for pre-season agriculture yield forecast." In *2018 IEEE 14th International Conference on e-Science (e-Science)*, pp. 423-430. IEEE, 2018.

[8] Kandasamy, Sathya Uma Lakshmi, Piyush Kumar Singh, and Dillip Kumar Swain. "Climate change vulnerability assessment of dryland farmers and factors Identification using machine learning techniques." *Local Environment* 27, no. 7 (2022): 824-846.

[9] Cai, Yaping, Kaiyu Guan, Jian Peng, Shaowen Wang, Christopher Seifert, Brian Wardlow, and Zhan Li. "A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach." *Remote sensing of environment* 210 (2018): 35-47

[10] Veronesi, Fabio, Athina Korfiati, René Buffat, and Martin Raubal. "Assessing accuracy and geographical transferability of machine learning algorithms for wind speed modeling." In *Societal Geo-innovation: Selected papers of the 20th AGILE conference on Geographic Information Science*, pp. 297-310. Springer International Publishing, 2017.

[11] Bilgilioğlu, Süleyman Sefa. "Land suitability assessment for Olive cultivation using GIS and multi-criteria decision-making in Mersin City, Turkey." *Arabian Journal of Geosciences* 14, no. 22 (2021): 2434.

[12] Gonzalez-Sanchez, Alberto, Juan Frausto-Solis, and Waldo Ojeda-Bustamante. "Predictive ability of machine learning methods for massive crop yield prediction." *Spanish Journal of Agricultural Research* 12, no. 2 (2014): 313-328.

[13] Ritz, Christian, E. Putku, and A. Astover. "A practical two-step approach for mixed model-based kriging, with an application to the prediction of soil organic carbon concentration." *European Journal of Soil Science* 66, no. 3 (2015): 548-554

[14] Parkes, Ben, Thomas P. Higginbottom, Koen Hufkens, Francisco Ceballos, Berber Kramer, and Timothy Foster. "Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change." *Environmental Research Letters* 14, no. 12 (2019): 124089.

[15] Ramirez-Villegas, Julian, and Andy Challinor. "Assessing relevant climate data for agricultural applications." *Agricultural and forest meteorology* 161 (2012): 26-45.

[16] Gopi, P. S. S., and M. Karthikeyan. "Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model." *Intelligent Automation & Soft Computing* 36, no. 1 (2023).

[17] Tamil Selvi, M., and B. Jaison. "Lemuria: a novel future crop prediction algorithm using data mining." *The Computer Journal* 65, no. 3 (2022): 655-666.

[18] Mallik, Santanu, AbhigyanChakraborty, Umesh Mishra, and Niladri Paul. "Prediction of irrigation water suitability using geospatial computing approach: a case study of Agartala city, India." *Environmental Science and Pollution Research* (2022): 1-16.

[19] Elavarasan, Dhivya, and P. M. Durai Raj Vincent. "Fuzzy deep learning-based crop yield prediction model for sustainable agronomical frameworks." *Neural Computing and Applications* 33, no. 20 (2021): 13205-13224.

[20] Khaki, Saeed, Hieu Pham, and Lizhi Wang. "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning." *Scientific Reports* 11, no. 1 (2021): 1-14.

[21] Mupangwa, W., L. Chipindu, I. Nyagumbo, S. Mkuhlani, and G. Sisito."Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa." *SN Applied Sciences* 2, no. 5 (2020): 1-14.

[22] Archana, K., and K. G. Saranya."Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier." *SSRG Int. J. Comput. Sci. Eng* 7 (2020): 1-4.

Copyrights @ Roman Science Publications Ins.

[23] Medar, Ramesh, Vijay S. Rajpurohit, and ShwetaShweta."Crop yield prediction using machine learning techniques." In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pp. 1-5. IEEE, 2019.

[24] Gopal, PS Maya, and R. Bhargavi. "A novel approach for efficient crop yield prediction." *Computers and Electronics in Agriculture* 165 (2019): 104968.

[25] Sun, Jie, Liping Di, Ziheng Sun, YonglinShen, and Zulong Lai. "County-level soybean yield prediction using deep CNN-LSTM model." *Sensors* 19, no. 20 (2019): 4363.

[26] Patil, Suraj, Varsha Nemade, and Piyush Kumar Soni. "Predictive modelling for credit card fraud detection using data analytics." *Procedia computer science* 132 (2018): 385-395.

[27] Hao, Lin, Juncheol Kim, Sookhee Kwon, and Il Do Ha. "Deep learning-based survival analysis for high-dimensional survival data." *Mathematics* 9, no. 11 (2021): 1244.

[28] Chen, Mingzhe, Ursula Challita, Walid Saad, Changchuan Yin, and Mérouane Debbah. "Artificial neural networks-based machine learning for wireless networks: A tutorial." *IEEE Communications Surveys & Tutorials* 21, no. 4 (2019): 3039-3071.

[29] Bre, Facundo, Juan M. Gimenez, and Víctor D. Fachinotti. "Prediction of wind pressure coefficients on building surfaces using artificial neural networks." *Energy and Buildings* 158 (2018): 1429-1441.

[30] Support Vector Machine [Online]. Available: https://www.fer.unizg.hr /_download/repository/07-SupportVectorMachine