REST Journal on Data Analytics and Artificial Intelligence
Vol: 3(2), June 2024
REST Publisher; ISSN: 2583-5564
Website: http: //restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/3/2/13

Wine Oracle: A Clandestine Journey into Quality Prediction with Neural Alchemy

* Winnish Allwin G J, Dhaanush K, Vuppu Reddy Kalyan, Mary Hanna Priyadharshini
Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College
Chennai, India
**Corresponding author: winnishallwin@gmail.com*

Abstract: *In the modern context, it is crucial to prioritize wine quality assurance in order to satisfy consumer demands and sustain a competitive advantage in the marketplace. The objective of this research is mainly focused on: (1) predicting the quality of wine through the utilization of sophisticated fused deep learning and machine learning methods namely Multichannel Convolutional Neural Network (MCNN). (2) Integrate machine learning models such as Multilayer Perceptron (MLP) to increase the model's prediction. These two methods exhibit a commendable degree of precision in forecasting the quality of wine. This study highlights the importance of utilizing advanced computational methods to improve quality control procedures, thus facilitating the wine sector's adjustment to changing consumer preferences and quality benchmarks.*

Keywords: *CNN, MLP, Deep learning, Neural Network, Classification, Prediction*

1. INTRODUCTION

The world's most popular libation is wine, and its qualities are held in the highest regard by society. The quality of wine is always essential for its consumers, particularly for producers in today's competitive market seeking to increase income. Historically, wine quality was calculated by testing after production; to attain that level, one already invested a significant quantity of time and money. If the quality is subpar, numerous procedures must be implemented from the start, which is quite expensive. Wine quality is vital since consuming low-quality wine is harmful to your health. The wine industry is enormous, and many consumers are misled about wine quality. Manually recognizing and classifying wine at various levels of quality is time-consuming and imprecise.

In this landscape, a series of research endeavors illuminate the path to a more efficient and accurate evaluation of wine quality. The initial research, using machine learning algorithms and synthetic data, focuses on predicting the quality of Pinot Noir wines from New Zealand. It accentuates the role of feature selection in enhancing classifier performance, particularly highlighting the superiority of an Adaptive Boosting classifier with crucial variables [1]. Transitioning to another facet, this research assesses wine quality based on particular parameters using machine-learning approaches such as genetic algorithms and simulated annealing. It stresses the significance of feature selection by demonstrating that sets based on feature selection may yield better predictions than sets based on all characteristics [2]. The third study takes a statistical approach to forecasting wine quality, examining the effectiveness of Ridge Regression, Support Vector Machine (SVM), Gradient Boosting Regress or, and Artificial Neural Network approaches. Gradient Boosting Regress or is identified as the best-performing model in the research [3]. Continuing the exploration in the fourth study, which examines consumer safety from an angle, classification algorithms for wine quality analysis—SVM, random forest, and multilayer perceptron—are compared. It emphasizes how important precise findings are to guarantee the quality and safety of food in the food sector [4]. Introducing a framework for predicting the quality of red wine utilizing MF-DCCA and predictive machine learning algorithms, the fifth study outperforms other commonly utilized algorithms, highlighting the significance of sophisticated modelling methods in determining wine quality [5]. Progressing to the sixth study - random forest, decision trees, and extreme gradient boosting are investigated for wine quality prediction; random forest is found to be the most exact and accurate model [6]. For wine quality prediction, the seventh paper presents a novel method that improves upon baseline approaches by combining PCA analysis, Pearson correlation analysis, and a 1D-CNN architecture [7]. Shifting focus, the eighth study emphasizes how important it is to assess red wine quality in terms of consumer health, using prediction methods like Naïve Bayes, Support Vector Machine, and Random Forest. To increase accuracy and efficiency, it proposes potential enhancement via feature extraction and combination [8].

2. LITERATURE SURVEY

Delving into the realm of wine quality prediction through machine learning, the initial dataset comprised 18 Pinot Noir wine samples, characterized by 54 different features. The Synthetic Minority Over-Sampling Technique (SMOTE) enriched the dataset to 1381 samples, with six reserved for model testing. The study explored seven machine learning models, including Adaptive Boosting (AdaBoost), Random Forest (RF), XGBOOST, and Stochastic Gradient Decision Classifier. Notably, the AdaBoost classifier exhibited remarkable performance, achieving 100% accuracy under various conditions. The primary goal was to predict wine quality by leveraging synthetic data and constructing a machine-learning model based on this synthesized data. The methodology involved comparing findings from four distinct feature selection approaches, utilizing essential variables identified in at least three methods to predict wine quality. [1] Moving to the prediction of wine quality through the application of machine learning algorithms, the second study emphasizes the analysis of multiple parameters crucial for determining wine quality. Although specific details about the dataset are not provided, the study explores Ridge Regression (RR), Support Vector Machine (SVM), Gradient Boosting Regress or (GBR), and a multi-layer Artificial Neural Network (ANN). Impressively, the Gradient Boosting Regress or (GBR) outperformed other models, boasting impressive Mean Squared Error (MSE), R, and Mean Absolute Percentage Error (MAPE) values. This research aims to predict wine quality based on various parameters, showcasing how statistical analysis can unveil the key components influencing wine quality before production. The methodology employed machine learning techniques to comprehend the data's structure. [2] Shifting focus to the prediction of wine quality through machine learning and hybrid modeling, the third study utilizes a dataset from the public repository of UCI machine learning. Key algorithms implemented include Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XG Boost). Although specific accuracy values are not provided, the paper emphasizes the utilization of accuracy, precision, recall, and F1 score for data interpretation. This research explores the role of machine learning techniques in hybrid models for enhancing wine quality prediction. The methodology involves leveraging machine learning and hybrid modeling techniques to predict wine quality. [3] The fourth study classifies wine using imbalanced data, employing a dataset of white wines from the Minho region in Portugal. The study utilizes the Synthetic Minority Over-Sampling Technique (SMOTE) to oversample the minority class and applies three different classification techniques: decision tree, adaptive boosting (AdaBoost), and random forest. Although specific accuracy values are not disclosed, the paper indicates that the random forest technique yields favorable results with the least percentage of error. The primary aim is to propose a data analysis approach for classifying wine into different quality categories. The balanced data is utilized to model a classifier categorizing wine into three categories: high quality, normal quality, and poor quality. The methodology involves employing the Synthetic Minority Over-Sampling Technique (SMOTE) and applying three different classification techniques. [4] The fifth study delves into predicting quality for different types of wine based on various feature sets, utilizing datasets encompassing both white and red wine quality. The study explores machine learning techniques to evaluate wine quality based on attributes indicative of quality. Different feature selection techniques, including genetic algorithm (GA)-based feature selection and simulated annealing (SA)-based feature selection, assess prediction performance. The study reports accuracy ranging from 95.23% to 98.81% with different feature sets, showcasing the efficacy of the proposed methodologies. The primary objective is to explore machine learning techniques for assessing wine quality based on attributes associated with quality. The methodology involves employing various feature selection techniques, including genetic algorithm (GA) and simulated annealing (SA), to evaluate prediction performance. [5] Introducing a generalized wine quality prediction framework employing evolutionary algorithms, the sixth study proposes a generalized wine quality prediction algorithm utilizing genetic algorithms. The unique approach involves encoding classifiers and their hyper parameters into a chromosome, evaluating the fitness based on the average accuracy of employed classifiers. Although specific accuracy values are not provided, the paper highlights the evaluation of fitness through the average accuracy of the classifiers. The primary aim is to propose a generalized framework for wine quality prediction, utilizing genetic algorithms to find a useful hybrid model. The methodology involves encoding classifiers and their hyper parameters into a chromosome, utilizing genetic operations to generate new offspring. [6]

The seventh study explores the realm of red wine quality prediction, employing advanced techniques such as One-dimensional Convolutional Neural Networks (1D-CNN). While the exact dataset is not specified, the study acknowledges the utilization of data mining algorithms on a wine quality dataset to analyze attributes like quality or class. The central algorithm, a 1D-CNN architecture, is incorporated to effectively capture correlations among neighboring features, enhancing the model's predictive capabilities. To bolster robustness, the study implements dropout and batch normalization techniques. Although specific accuracy values are not provided, the paper underscores the superior performance of their proposed method through extensive experiments, outperforming baseline approaches in wine quality prediction. The primary objective is to apply machine learning techniques to analyze attributes within the wine quality dataset, complemented by Pearson correlation analysis, PCA analysis, and Shapiro-Wilk tests on these properties. This comprehensive methodology highlights the study's commitment to a multifaceted approach for red wine quality prediction. [7] This study introduces a novel framework for red wine quality prediction, leveraging machine learning techniques. The dataset, sourced from the UC Irvine Machine Learning Repository, comprises 1599 instances of red wine, featuring 11 physicochemical data attributes. The study utilizes the Multifractal Detruded Cross-Correlation Analysis (MF-DCCA) method to quantitatively investigate cross-correlations between quality and physicochemical data. Additionally, the research compares the performance of two machine learning algorithms against other common models implemented on the red wine dataset. While specific accuracy values are not provided, the paper asserts that their model demonstrates superior performance compared to alternative approaches. The primary aim is to propose an innovative framework for predicting red wine quality ratings, with a focus on enhancing prediction efficiency and accuracy. The methodology encompasses using MF-DCCA to quantitatively analyze cross-correlations and rank the importance of long-range correlations. This approach underscores the study's commitment to advancing the precision and effectiveness of red wine quality prediction. [8]

3. DATASET

In the pursuit of our research endeavors, the selection of an appropriate dataset plays a pivotal role in shaping the outcomes of our predictive model. The axiom “Garbage in, Garbage Out”, succinctly underscores the critical connection between the quality of the dataset used as input and the readability of the ensuing results. Our research focuses on a dataset pertaining to the quality of the red wine samples, representing a comprehensive repository of information. This dataset encompasses a multitude of numerical attributes that capture diverse chemical characteristics inherent in red wines. Being presented with 4086 examples, each representing a distinct red wine sample, the size of the data set is carefully calibrated to achieve a delicate balance. This delicate balance is designed to mitigate any computational limits while also providing enough data for reliable model training. A number of prominent characteristics of the dataset provide insights into different aspects of the composition and characteristics of red wine.

The key attributes encapsulated within the dataset are:

- **Fixed Acidity:** Denoting the non-volatile acids present in the wine.
- **Volatile Acidity:** Representing the amount of acetic acid, which can impart an undesirable vinegar taste.
- **Citric Acid:** Signifying the concentration of citric acid, contributing to freshness and flavor.
- **Residual Sugar:** The amount of residual sugar remaining after fermentation.
- **Chlorides:** The concentration of salts in the wine.
- **Free Sulfur Dioxide**: The free form of SO2, which prevents microbial growth and oxidation.
- **Total Sulfur Dioxide:** The total amount of SO2, including both free and bound forms.
- **Density:** The density of the wine.
- **pH:** The acidity or basicity of the wine.
- **Sulphates:** The amount of sulfates in the wine, contributing to its antimicrobial properties.
- **Alcohol:** The alcohol content of the wine

In elucidating the comprehensive particulars of the Red and White Wine Dataset, the subsequent delineation encapsulates the distinctive attributes of both datasets.

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide
count	1599.0	1599.0	1599.0	1599.0	1599.0	1599.0
mean	8.319637273295838	0.5278205128205131	0.2709756097560964	2.5388055034396517	0.08746654158849257	15.874921826141339
std	1.7410963181277006	0.17905970415353498	0.19480113740531785	1.4099280595072805	0.047065302010090154	10.46015696980973
min	4.6	0.12	0.0	0.9	0.012	1.0
25%	7.1	0.39	0.09	1.9	0.07	7.0
50%	7.9	0.52	0.26	2.2	0.079	14.0
75%	9.2	0.64	0.42	2.6	0.09	21.0
max	15.9	1.58	1.0	15.5	0.611	72.0

FIGURE 1. Description of Red Wine of Characteristics
(Fixed Acidity, Volatile Acidity, Citric Acid, Residual Sugar, Chlorides, Free Sulfur Dioxide)

total sulfur dioxide	density	pH	sulphates	alcohol	quality
1599.0	1599.0	1599.0	1599.0	1599.0	1599.0
46.46779237023139	0.9967466791744831	3.311113195747343	0.6581488430268921	10.422983114446502	5.6360225140712945
32.89532447829901	0.0018873339538425559	0.15438646490354266	0.16950697959010977	1.0656675818473926	0.8075694397347023
6.0	0.99007	2.74	0.33	8.4	3.0
22.0	0.9956	3.21	0.55	9.5	5.0
38.0	0.99675	3.31	0.62	10.2	6.0
62.0	0.997835	3.4	0.73	11.1	6.0
289.0	1.00369	4.01	2.0	14.9	8.0

FIGURE 2. Description of Red Wine Characteristics
(Total Sulfur Dioxide, Density, pH, Sulphates, Alcohol, Quality)

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide
count	4898.0	4898.0	4898.0	4898.0	4898.0	4898.0
mean	6.854787668436075	0.27824111882401087	0.33419150673743736	6.391414863209486	0.0457723560636995	35.30808493262556
std	0.8438682276875188	0.10079454842486428	0.12101980420298301	5.072057784014864	0.02184796809372882	17.007137325232566
min	3.8	0.08	0.0	0.6	0.009	2.0
25%	6.3	0.21	0.27	1.7	0.036	23.0
50%	6.8	0.26	0.32	5.2	0.043	34.0
75%	7.3	0.32	0.39	9.9	0.05	46.0
max	14.2	1.1	1.66	65.8	0.346	289.0

FIGURE 3. Description of White Wine Characteristics
(Fixed Acidity, Volatile Acidity, Citric Acid, Residual Sugar, Chlorides, Free Sulfur Dioxide)

total sulfur dioxide	density	pH	sulphates	alcohol	quality
4898.0	4898.0	4898.0	4898.0	4898.0	4898.0
138.36065741118824	0.9940273764801896	3.1882666394446693	0.4898468762760325	10.514267047774638	5.87790935075541
42.49806455414294	0.0029909069169369354	0.1510005996150667	0.11412583394883138	1.2306205677573183	0.8856385749678454
9.0	0.98711	2.72	0.22	8.0	3.0
108.0	0.9917225000000001	3.09	0.41	9.5	5.0
134.0	0.99374	3.18	0.47	10.4	6.0
167.0	0.9961	3.28	0.55	11.4	6.0
440.0	1.03898	3.82	1.08	14.2	9.0

FIGURE 4. Description of White Wine Characteristics
(Total Sulfur Dioxide, Density, pH, Sulphates, Alcohol, Quality)

For enhanced clarity and comprehensibility, the ensuing histograms depict the physiochemical properties of Red Wine and White Wine in graphical form.

Histogram of Volatile Acidity:

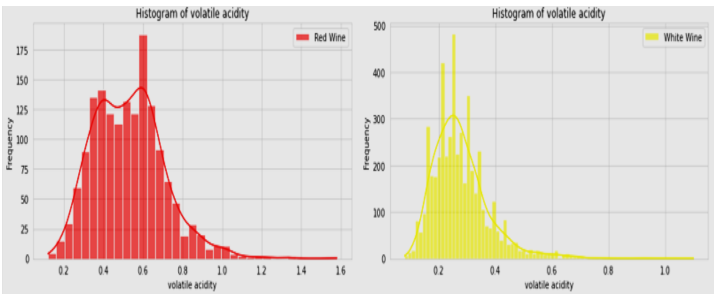

FIGURE 5. Red wine vs White Wine

Histogram of Citric Acid:

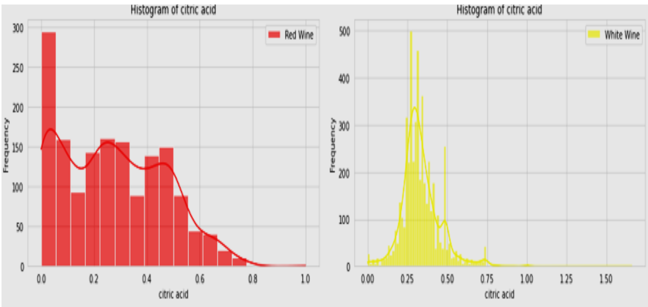

FIGURE 6. Red Wine Vs White Wine

Histogram of Residual Sugar:

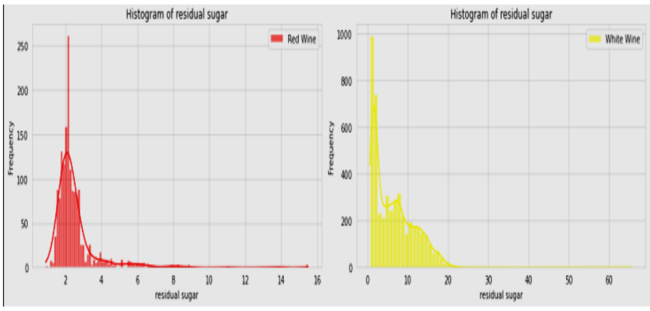

FIGURE 7. Red Wine Vs White Wine

Histogram of Chlorides:

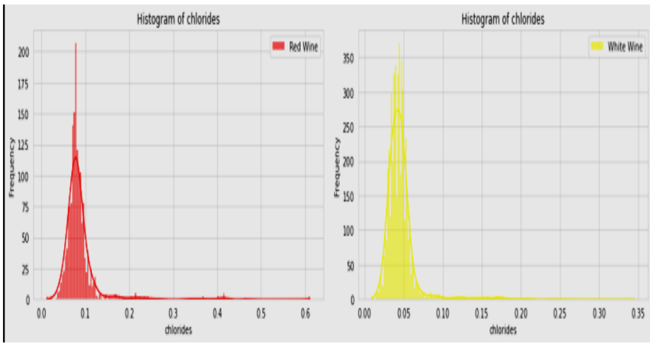

FIGURE 8. Red Wine Vs White Wine

Histogram of Free Sulfur Dioxide:

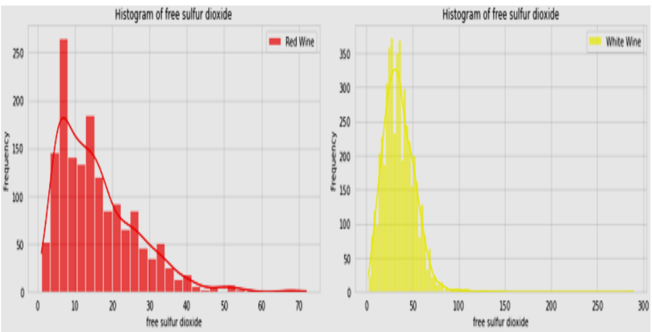

FIGURE 9. Red Wine Vs White Wine

Histogram of Density:

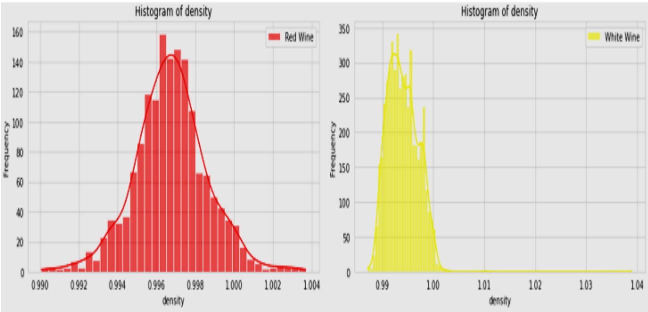

FIGURE 10. Red Wine Vs White Wine

Histogram of Ph:

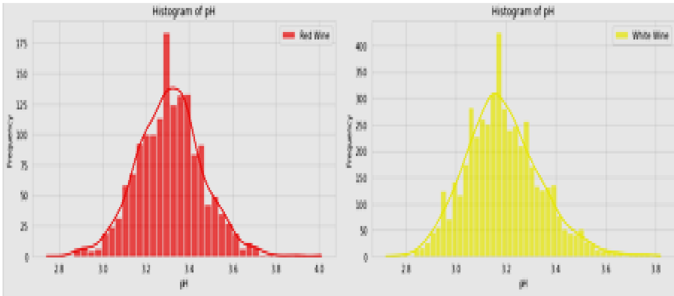

FIGURE 11. Red Wine Vs White Wine

Histogram of Sulphates:

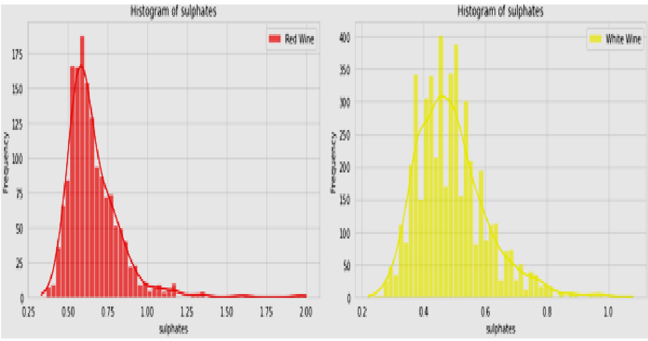

FIGURE 12. Red Wine Vs White Wine

Histogram of Alcohol:

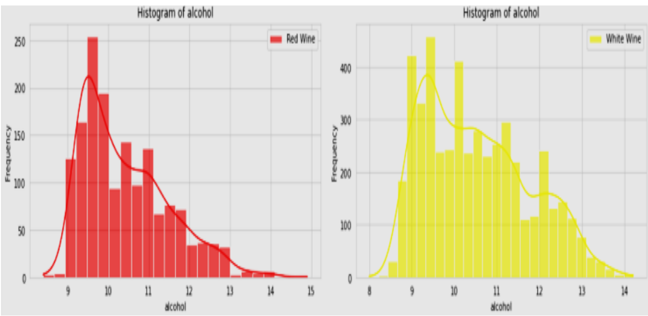

FIGURE 13. Red Wine Vs White Wine

4. DATA LOADING AND PREPROCESSING

After loading the dataset (wine_quality_red_balanced.csv), duplicates are eliminated.
Engineering Features: We round the target variable, quality, to the closest whole value. A feature's 'X' (input features) and 'y' (target) are different.
Scaling Data: Standard Scaler is used to apply standard scaling to the input characteristics. By eliminating the mean and scaling to the unit variance, it standardizes the features.
Reshaping for CNN: To make input features compatible with Conv1D layers, they are altered. In order to represent the channels in a third dimension, one channel for each feature - the channels are reshaped.
CNN One-Hot Encoding: For CNN, the target variable 'quality' is one-hot encoded. Binary vectors are created from category data using one-hot encoding.
To make input features compatible with Conv1D layers, they are altered. In order to represent the channels in a third dimension—one channel for each feature—the channels are reshaped.

5. PROPOSED MODEL

In our research, we employ a sophisticated model architecture that combines the strengths of a multichannel Convolutional Neural Network (CNN) and a Multilayer Perceptron (MLP) to enhance the predictive capabilities for wine quality assessment.

Multichannel Convolutional Neural Network (MC-CNN)
A Multichannel Convolutional Neural Network is a type of deep learning model that's used for text classification and sentiment analysis. It's an extension of the standard CNN model, which used a word embedding layer and one-dimensional convolutional neural network for text classification.
The CNN component of our model leverages a multichannel architecture, designed to process different subsets of input features independently. This approach aims to capture complex patterns and relationships within the data. The multichannel design is implemented as follows:
Input Channels:
The input features are split into two channels, allowing the model to extract distinct patterns from each subset. Channel 1 processes the first half of the input features, while Channel 2 processes the remaining features.

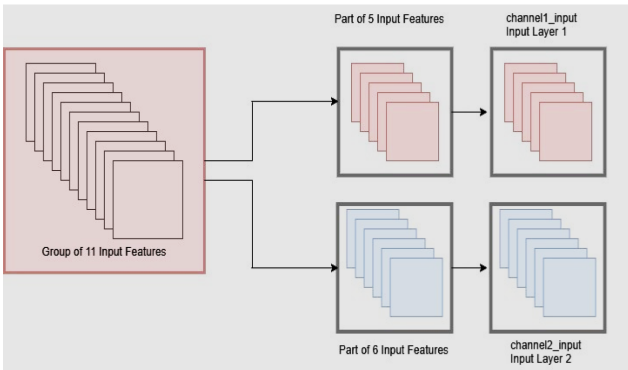

FIGURE 14. Separation of Input Features

Hidden Layer: The hidden layer receives the input that was previously given into the input layer. Several hidden layers may exist, contingent on the amount of the data and our model. The number of neurons in each hidden layer varies, but they are usually more than the number of characteristics.

Convolutional Layers: Two convolutional layers are applied to each channel independently, each comprising 32 filters with a kernel size of 3 and ReLU activation.
MaxPooling layers follow the convolutional layers, providing spatial down-sampling to capture essential information.
Flattening and Concatenation:
The output from each channel is flattened and then concatenated to capture the combined feature representation.

Dense Layers:
Two dense layers with 128 and 64 units, respectively, and ReLU activation are employed for further feature processing.
Dropout layers (with a dropout rate of 0.5) are incorporated to prevent overfitting.

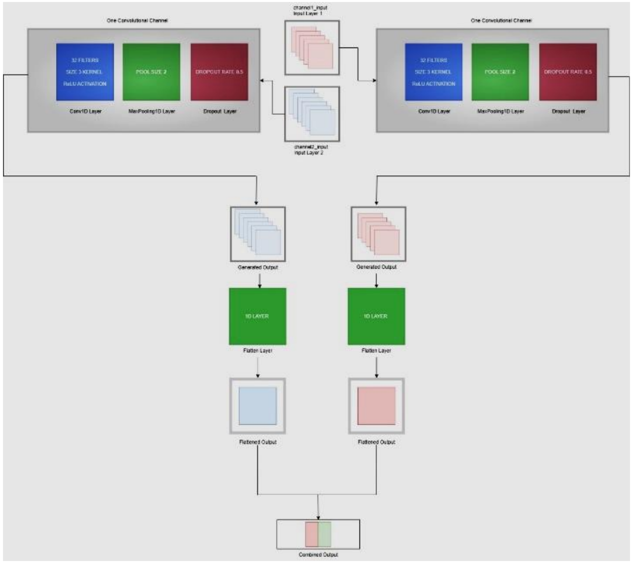

FIGURE 15. Outcome of Combined Output

Output Layer: The output layer utilizes the softmax activation function, enabling the model to predict wine quality categories**.**

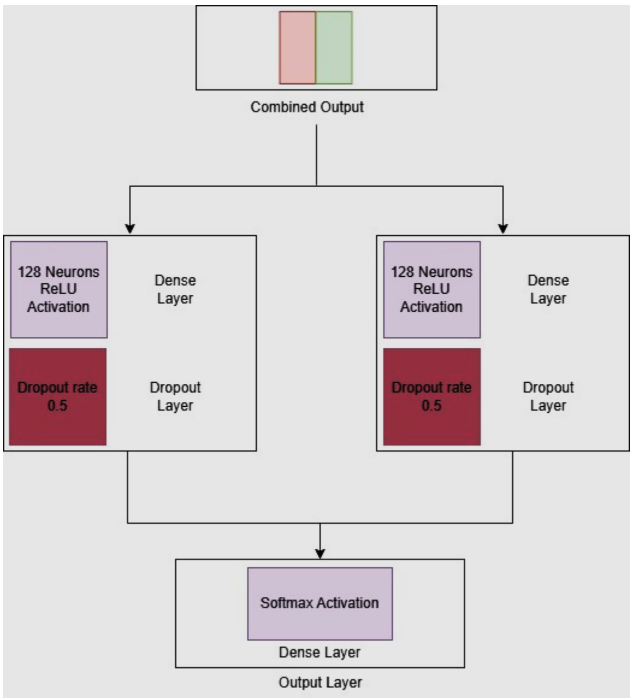

FIGURE 16. Generation of Output Layer

Multilayer Perceptron (MLP): MLP stands for multi-layer perception. Dense layers that are entirely linked convert any input dimension to the required dimension. A neural network with several layers is called a multi-layer perception. Neurons are joined together to form neural networks, with some neurons' outputs acting as inputs for other neurons. In multi-layer perception, the number of nodes in each hidden layer and the total number of hidden layers are both infinite. Single neurons, or nodes, comprise each input, output, and hidden layer; for each output, there are numerous nodes in the hidden layer, and so on.

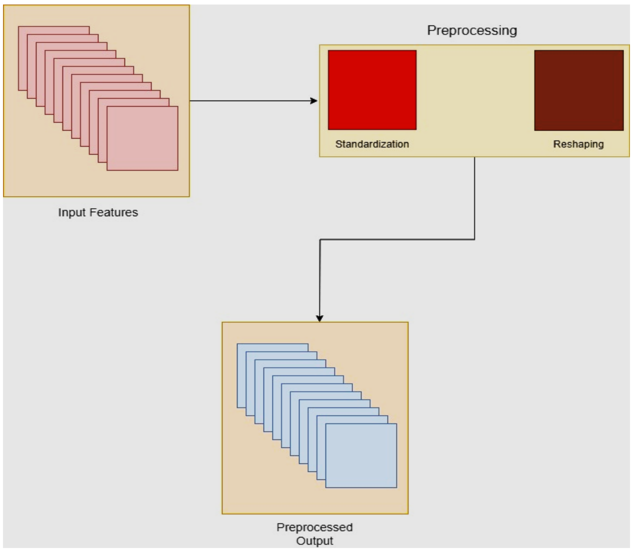

FIGURE 17. Generation of Preprocessed Output

Hidden Layers: The MLP comprises two hidden layers with 64 and 32 units, respectively, fostering nonlinear transformations.

Output Layer: Similar to the CNN, the MLP output layer employs the softmax activation function for categorical prediction.

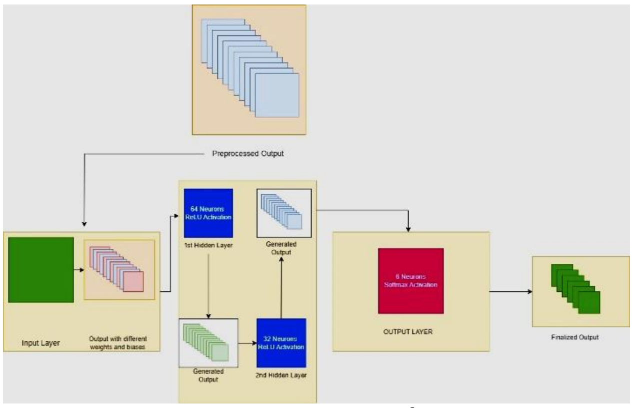

FIGURE 18. Outcome of Output

Combining MC-CNN and MLP: The outputs of the CNN and MLP are concatenated, creating an integrated model that captures intricate relationships within the data. This multifaceted architecture, blending the strengths of a multichannel CNN and an MLP, enables our model to discern nuanced patterns in the input data, resulting in a robust and accurate predictor for wine quality.

Model Architecture:
Input Channels:
Two channels from the CNN model and one channel from the MLP model.
Dense Layers:
Two dense layers with ReLU activation and dropout.
Output Layer:
Dense layer with softmax activation.
Mathematical Representation:

Let's denote:
Xcnn: Input to the CNN model.
Xmlp: Input to the MLP model.
ycnn: Output from the CNN model.
ymlp: Output from the MLP model.
ycombined: Output from the combined model.
The mathematical representation can be written as:
CNN Model: ycnn=CNN(Xcnn)
MLP Model: ymlp=MLP(Xmlp)

Combined Model:

ycombined=Combined ([ycnn, ymlp])

6. CONCLUSION

This study aimed to enhance wine quality prediction using a self-generated dataset of 4086 red wine instances, characterized by 11 physicochemical properties. We utilized a hybrid model combining a Multichannel Convolutional Neural Network (MC-CNN) with a Multilayer Perceptron (MLP). The MC-CNN processed feature subsets independently, while the MLP integrated these features for improved predictions. Techniques like dropout and batch normalization were used to enhance model robustness and accuracy.

Our evaluation showed that the MC-CNN achieved 60.5% of accuracy, the MLP reached 87.5% of accuracy, and their combination significantly improved the total accuracy to 87.6% for the red wine dataset and 90% for the white wine dataset. This demonstrates the effectiveness of our hybrid approach over traditional models. In summary, our integration of advanced machine learning techniques with careful data preprocessing resulted in a robust framework for predicting wine quality, offering a promising foundation for future research.

REFERENCES

[1]. Terry Hui-Ye Chiu, Chienwen Wu, Chun-Hao Chen, A Generalized Wine Quality Prediction Framework by Evolutionary Algorithms, Vol. 6, No.7, Taipei, 106 (Taiwan), Department of Information and Finance Management, National Taipei University Technology, .2021.

[2]. Piyush Bhardwaj, Parul Tiwari, Kenneth Olejar Jr, Wendy Parr, Don Kulasiri, A machine learning application in wine quality prediction, Volume 8, New Zealand, Elsevier Ltd, 2022.

[3]. K. R. Dahal, J. N. Dahal, H. Banjade, S. Gaire, Prediction of Wine Quality Using Machine Learning Algorithms, Vol.11 No.2, USA, Open Journal of Statistics, 2021

[4]. Kanika Agrawal, Nelshan Mandan, Sunny Kumar, Red Wine Quality Prediction Using Machine Learning Techniques, India, IEEE, 2020.

[5]. Chao Ye, Ke Li and Guo-zhu Jia, A New Red Wine Prediction Framework Using Machine Learning, China, IOP Publishing Ltd, 2020.

[6]. Devika Pawar, Aakanksha Mahajan, Sachin Bhoithe, Wine Quality Prediction using Machine Learning Algorithms, Volume 8–Issue 09, India, International Journal of Computer Applications Technology and Research, 2019

[7]. Bhavya AG, Wine Quality Prediction Using Different Machine Learning Techniques, India, International Journal of Science, Engineering and Technology, 2020.

[8]. Bipul Shaw, Ankur Kumar Suman, Biswarup Chakraborty, Wine Quality Analysis Using Machine Learning, India, Springer Nature Singapore Pte Ltd, 2020.

[9]. Avinash Sanjay Gawale, Wine Quality Prediction using Machine Learning and Hybrid Modeling, India, National College of Ireland, 2022.

[10]. A. Trivedi and R. Sehrawat, "Wine Quality Detection through Machine Learning Algorithms," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018.

[11]. Research on Wine Quality Prediction Based on Support Vector Machine Method. Hans Journal of Data Mining. (2016). 06. 42-53. 10.12677/HJDM.2016.61006.

[12]. Siphendulwe Zaza, Marcellin Atemkeng, Sisipho Hamlomo, “Wine feature importance and quality prediction: A comparative study of machine learning algorithms with unbalanced data”, arXiv, 2023.

[13]. S. Di and Y. Yang, “Prediction of Red Wine Quality Using One-dimensional Convolutional Neural Networks”, arXiv, 2023.

[14]. Chao Ye, Ke Li and Guo-zhu Jia, “A new red wine prediction framework using machine learning”, Journal of Physics, 2020.

