



# A Study of Modeling Techniques for Prediction of Wine Quality

Ashley Laughter<sup>(✉)</sup> and Safwan Omari

Lewis University, Romeoville, IL 60446, USA

{ashleyrlaughter, omarisa}@lewisu.edu

**Abstract.** The evaluation of wine quality is multifactorial and involves scoring characteristics that relate to a wine's appearance, aroma and taste. These scores are, at least in part, subjective to individual expertise, experience and interpretation. Given this constraint, it is not surprising that the development of an objective, scientifically-based system for quantitatively scoring wine has remained elusive. In this work, multiple classification-based systems for predicting wine quality based on a set of measured physicochemical properties in the final product are evaluated. Several data pre-processing techniques are employed including, z-score normalization, outlier reduction, and principal component analysis (PCA). The resulting transformed data is then applied to various machine learning models (i.e., Multilayer Neural Networks (MNN), Random Forest (RF), basic Decision Tree (DT) and Support Vector Machines (SVM)). Standard metrics (i.e., confusion matrix, classification accuracies, precision, recall, F1-score), are compared for each model. A comparison of the results obtained in this study to contemporary works using similar metrics shows that two of the classifiers used in the current study (i.e., MNN with Adam optimizer and Random Forest) performed with higher classification accuracies and precision (70%/73 and 72%/73, respectively).

**Keywords:** Wine quality · Principle components · Classification · Dimensionality reduction · Normalization · Outlier · Scaling · Machine learning · Decision tree · Random forest · Multilayer neural network · Support vector machine

## 1 Introduction

Wine is one of the most popular choices for a social alcoholic beverage available today and is manufactured and consumed on a global scale. Therefore, wine as a commodity is both socially and economically important. From an economic perspective, the business of wine-making can be a highly profitable venture with the profit margin being closely tied to the overall quality of the wine being produced in any given area. The main factors that contribute to a wine's quality are governed by the tester's perception and experience at the time of tasting [1]. The evaluation of wine quality is considered *multifactorial*, with many different variables affecting the final rating. At least in part, the evaluation of wine quality encompasses a subjective domain and is intrinsically related to an individual's

taste preferences and cultural background along with other attributes. General factors that affect wine quality are the geographical origin and age (vintage), as well as the absence of common wine defects such as high amounts of acetic acid from bacterial contamination, cork taint due to oxidation by-products and the presence of precipitates from insufficient filtering and fining in the final product [1].

Wine evaluation can be sub-categorized into physicochemical and sensory-based tests as reported by Cortez et al. [2], Appalasamy et al. [3], and Ebeler [4]. The latter method of evaluation relies on the skill and experience of trained human experts, i.e. sommeliers, skilled vintners, professional wine tasters, etc. During the sensory evaluation, the subjective sensations of taste and smell are routinely used to score wines and thus provide a final quality rating (usually ordered numerical ratings) for the specific wine under study. Physicochemical testing is performed using standard laboratory methods to quantitatively measure the levels of various elements inherent to a specific wine, such as alcohol percentage, density, sulphates, pH, etc. The focus of this study is to evaluate multiple classification-based modelling systems for predicting wine quality based on measured physicochemical properties.

In the current study, variants of the Portuguese “vinho verde” wine are analyzed. The ‘Wine Quality Dataset’ can be found in the UCI Irvine Machine Learning Repository [5]. The data were collected from May of 2004 to February of 2007 using the protected designation of origin samples tested at the official certification entity (i.e., Comissão de Viticultura da Região dos Vinhos Verdes or CVRVV) [2]. The data consists of two separate datasets, one for red wine (consisting of 1,599 instances) and one for white wine (consisting of 4,898 instances). Both datasets contain eleven attributes based on routinely performed laboratory physicochemical tests. The chemical attributes are fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates and alcohol percentage. To generate this data, each wine sample was evaluated based on a blind sensory assessment using a minimum of three human assessors (wine experts); the final sensory score represents the quality rating given by calculating the median of the three scores [12]. The quality scale used to grade the samples ranged from a low score of ‘0’ (very bad) to the highest score of ‘10’ (excellent). Note that for both datasets (red/white), no samples received quality scores from 0 to 2 or 10 for any of the available instances. Additionally, for the red wine dataset (not used in this work), no samples received a score of 9.

The rest of the paper is organized as follows; related works are discussed in Section II, raw dataset features and preprocessing steps are detailed in Sect. 3, various classification models are reviewed, deployed and evaluated in Sect. 4, model results are reported and highlighted in Sect. 5, overall findings are discussed and reviewed in Sect. 6, the paper is concluded in Sect. 7 and future work and other recommendations are outlined in Sect. 8.

## 2 Related Work

A number of independent analyses have been performed using the ‘Wine Quality’ dataset available from the UCI Machine Learning repository [5]. Cortez et al. (2009) used a novel regression method to perform simultaneous variable and model selection for neural networks (NN), multiple regression (MR), and support vector machines (SVM) [2]. The

authors focused on predicting the quality of the wine based on the physicochemical properties, or attributes of the dataset. They concluded the better performance of the SVM (64.6% accuracy) for this task may be due to the differences in the training phase. The SVM guarantees an optimum fit whereas the NN may fall into a local minimum. Additionally, the SVM gives a linear penalty to large errors and is less sensitive to outliers while the NN minimizes the sum of squared errors. With this in mind, the model could be used for predicting the quality of wine and help in training oenology students.

In a second study conducted by Cortez et al. (2009), the authors focused on applying a data mining approach to the prediction of wine preferences using the Wine Quality dataset [12]. Similar to their previous work [2], the authors chose a regression approach used in combination with feature and model selection. Relative input importance was calculated for each of the eleven variables using the SVM model and showed that alcohol %, sulphates, pH, free sulfur dioxide and volatile acidity were the top five most important features. For this work, the support vector machine (SVM) outperformed multiple regression and neural network methods. Precision for each individual quality class (classes 4 to 8 only) ranged from 60.1% for class 6 to 86.6% for class 8 at a tolerance of 0.5. Classes 3 (poor) and 9 (excellent) were not included as these classes did not contain any entries in the matrix. Note that the overall model accuracy based on the confusion matrix data was 64.3% for the SVM model at a tolerance of 0.5.

Unlike Cortez et al. [2, 12], Appalasamy et al. (2012), focused on classification, rather than regression, to predict the quality of the wines based on their physicochemical properties [3]. Similar to the current study, Appalasamy et al., used data pre-processing prior to modeling. Specifically, the authors chose to discretize the numeric attributes to nominal values and also reduce the dimensionality of the attribute set. The results of the pre-processing therefore reduced the full attribute set to six attributes for the white wine data (i.e., volatile acidity, citric acid, chlorides, free sulfur dioxide, density and alcohol %) and four attributes for the red wine data (i.e., volatile acidity, total sulfur dioxide, sulphates, and alcohol %). The resulting data sets were then modeled using Decision Tree (DT) and Naïve Bayes (NB) classifiers. Appalasamy et al., concluded that the DT, specifically ID3 (Iterative Dichotomiser 3), outperformed the NB model when applied to the red wine data and resulted in an overall accuracy of 60.0% as compared with 58.8%. For the white wine data, an overall accuracy of 52.3% as compared with 50.5% was achieved. Results of the feature selection performed by Appalasamy et al., showed that alcohol% is the primary attribute across both data sets that provides the most information to determine wine quality. Like alcohol%, volatile acidity was ranked highly during the feature selection process and is known to also contribute to the evaluation of wine quality. In specific, volatile acidity relates to acetic acid content in wine and is therefore a negative discriminator; higher volatile acidity results in lower quality scores. In both [2] and [3], alcohol percentage and volatile acidity were shown to have the strongest impact on wine quality.

A study performed by Hu et al. (2016), addressed the quality class imbalance associated with the Wine Quality dataset during data pre-processing using the Synthetic Minority Over-Sampling (SMOTE) technique [13]. As Hu et al., observe, quality classes at the extremes (3, 4 and 9, 10) are underrepresented in the data set with the majority of class instances belonging to quality classes 5 and 6 (average or above average quality

wines). To address class imbalance, two solutions are commonly employed, either random undersampling of the majority classes to more closely match the minority classes, or, as in this study, over-sampling the minority classes by creating synthetic data derived directly from these classes (i.e., SMOTE [18]). Hu et al., note that SMOTE can replicate instances of the minority classes by using interpolation to increase the number of instances. Similar to the current study, Hu et al., reduced the number of classes from seven to three and assigned classes 3 and 4 to a ‘low’ quality group, 5, 6 and 7 to a ‘normal’ quality group, and 8, 9 to a ‘high’ quality group; this design therefore resulted in a 3 X 3 confusion matrix. Overall classification accuracies obtained from Hu et al., after using SMOTE pre-processing were 92% (Decision Tree), and 93.7% (SVM). The primary contribution of this study showed that SMOTE is effective at remediating class imbalance within a data set and can greatly improve classification accuracy of standard models (i.e., SVM, RF, and DT). As in Cortez et al. [2], Hu et al., performed feature importance analysis using the criterion of mean decrease in Gini values for the eleven attributes and showed that volatile acidity, free sulfur dioxide, alcohol %, citric acid and pH contributed the most information to the quality determination.

Er et al., (2016) used SVM, k-nearest neighbors (KNN) and random forest (RF) classifiers on the full Wine Quality data set (both red and white wines together) to classify each instance as red or white respectively [14]. All classifiers scored above 99.0% precision and recall showing that the two wine types can be accurately and efficiently separated using the full attribute set along with a standard classifier algorithm. Er et al., also used SVM, KNN and RF models to classify the quality of both data sets before and after principle component analysis transformation (PCA). All eleven quality groups (i.e., 0–10) were used for both sets of analyses (red and white wines) which differs from the current study in that the quality groups were condensed to four classes. Best recall accuracies reported by Er et al., prior to PCA transformation for the white wine data set were 70.3% using cross validation and 68.7% using percentage split, respectively, for the RF classifier model. Recall accuracies achieved post-PCA transformation for the white wine data set were 69.9% and 67.5%, using the RF classifier model. Comparable recall (and precision) accuracies were obtained for the red wine dataset. The authors note that higher recall (and precision) metrics were obtained after PCA transformation for the red wine data set; however, slightly lower accuracies were obtained for the white wine data set after PCA. The study performed by Er et al., is similar to the current study in that PCA transformation was applied to the data set prior to modeling.

In another related study, Gupta (2018) used linear regression to determine the feature importance for each of the eleven physicochemical attributes from the Wine Quality data set [15]. Each attribute (alcohol %, fixed acidity, volatile acidity, etc.) was therefore treated as an independent variable and regression coefficients calculated along with corresponding p-values using the quality score as the dependent variable. The resulting regression summary showed individual attribute contribution to the quality rating for both the red and white wine data sets. Attributes displaying the highest regression coefficients and lowest p-values were chosen for the downstream classifiers. For the white wine data set, the regression summary selected an eight feature subset consisting of fixed acidity, volatile acidity, residual sugar, free sulfur dioxide, density, pH, sulphates and alcohol% as the primary quality predictors. Similar to two previous studies by Cortez et al., [2, 12],

Gupta trained neural network and SVM regression models to predict the quality rating of the wine samples from both data sets using all eleven features and also the reduced feature sets. For both models, the reduced feature sets resulted in lower training, testing and validation errors. The confusion matrices, along with precision, recall and F1 score metrics were not reported. The primary contribution of this study showed that reducing the feature set by performing regression based feature selection for the Wine Quality data can result in lower overall prediction error.

In a related work, Aich et al., (2018), used both genetic algorithm (GA) and simulated annealing (SA) approaches to implement feature based selection for the Wine Quality data set [16]. Multiple classifiers (i.e., RPART, C4.5, RF, SVM, etc.) were used to make wine quality predictions on each of the reduced features sets. Aich et al., report high classification accuracies for most algorithms used (i.e., 95.2 to 98.8%), but do not show the raw data output for the corresponding confusion matrices, feature selection analysis and output for the GA or SA algorithms, or exactly which attributes were used in the reduced feature sets applied to the classification algorithms.

A second related study conducted by Aich et al., (2019) used a similar approach as in [16] and implemented an RFE (Recursive Feature Elimination) algorithm to produce a reduced feature set from the Wine Quality data set prior to classification [17]. RFE is a recursive process that provides a rank ordering of all features according to importance based on a selection criterion such as IG (Information Gain). As in [16], multiple classifiers are used to provide quality predictions on both the reduced and full feature sets and high accuracies are reported (i.e., 97.6 to 99.6%); however, confusion matrices, feature selection analysis details and exact attributes comprising the reduced feature sets were not shown. Additionally, no pre-processing (scaling, normalization, outlier detection, PCA, etc.) steps are indicated in either [16] or [17]. An examination of the Wine Quality data set reveals significant asymmetry with respect to quality class distribution, and disparity with respect to individual attribute scale. These issues are known to affect classification accuracy but were not addressed in either [16] or [17].

In the current study, we build on existing work by conducting additional preprocessing steps applied to the dataset to adjust the scale and remedy the high variance instances for some attributes. Additionally, multiple classification models are developed to improve overall accuracy metrics as compared to related works. Initial data exploration determined that most of the eleven feature attributes contain high variance instances; such data points can severely impact classification accuracy, therefore, outlier removal is implemented using z-score normalization, and instances exceeding  $\pm 3$  standard deviations from their respective means are dropped. Given that the data are normally distributed, removal of instances exceeding three standard deviations from the mean is therefore recommended as these points are likely outliers.

Furthermore, significant class imbalance exists within the dataset at the quality extremes (low and high) such that average quality instances are overrepresented. Initial ANOVA analysis of quality group means for each attribute showed that some quality groups do not significantly differ with respect to their means for each attribute at alpha = 0.05 (data not shown). Based on this initial assessment, the quality classes were therefore condensed from seven (3–9) to four (4–7) to provide increased discriminatory power for the downstream classification models. To the best of our knowledge, this study is the

first to implement an outlier removal strategy prior to downstream analyses. After data pre-processing, multilayer neural networks, random forest, support vector machines and decision tree models are used to predict the quality classification for each instance.

### 3 Methodology

In this study, Portuguese '*vinho verde*' wine data is pre-processed and various classification modeling systems are built. The goal is to predict wine quality based on the physicochemical properties. The 'Wine Quality Dataset' is publicly available in the UCI Irvine Machine Learning Repository [5]. It consists of two separate datasets – one for red wine (consisting of 1,599 instances) and one for white wine (consisting of 4,898 instances). Both datasets contain eleven attributes capturing a specific wine sample physicochemical property, as well as a quality score between 0 and 9, with a score of '0' representing lowest wine sample quality. The available attributes are: (1) Fixed Acidity (tart, sourness), (2) Volatile Acidity (volatile organic acids, i.e., acetic acid, spoilage), (3) Citric Acid (nonvolatile acid, freshness), (4) Residual Sugar (sugar remaining after fermentation), (5) Chlorides (salt content), (6) Free Sulfur Dioxide (free SO<sub>2</sub>, antimicrobial agent), (7) Total Sulfur Dioxide (amount of free and bound SO<sub>2</sub>), (8) Density (density of water based on alcohol and sugar levels), (9) pH (level of alkalinity or acidity), (10) Sulphates (wine additive, contributes to SO<sub>2</sub> gas levels) and (11) Alcohol percentage (% of alcohol present). The final output is a quality score, which is a numeric, ordered value. The dataset can be viewed as a classification or regression task but may also be utilized in unsupervised learning as well. Due to its smaller size (<1,600 instances), and similar group distribution to the white wine data set, the red wine dataset was therefore not used in this study.

During initial data set exploration, we observed high variance observations and unequal distribution of samples in the quality classes. These issues are predicted to negatively impact the performance of the downstream models, especially the multilayer neural network (i.e., MNN). As a result, multiple preprocessing steps were performed, including: z-score normalization of data, outlier removal, re-labeling and coding of class samples, quality class reduction and condensation, and finally, Principal Component Analysis (PCA) as detailed in the following sections.

#### 3.1 Normalization and Scaling

For the Wine Quality data set, all of the eleven feature attributes are numerical; however, the values exhibit a large disparity with respect to scale. Fixed acidity ranges from 3.8 to 14.2, whereas, residual sugar ranges from 0.6 to 65.8. The following Table 1 presents the data statistics for a few selected features from the total attribute set.

Based on these statistics, we concluded that initial data processing (z-score normalization) was required in order to place all data points on the same scale with mean of 0 and standard deviation of 1, and also to assist with eventual downstream analyses, in particular, the multilayer neural network models. To accomplish this task, the python programming language [6] was used, more specifically, the Spyder integrated development environment (IDE) from the Anaconda distribution [7]. For the normalization, all

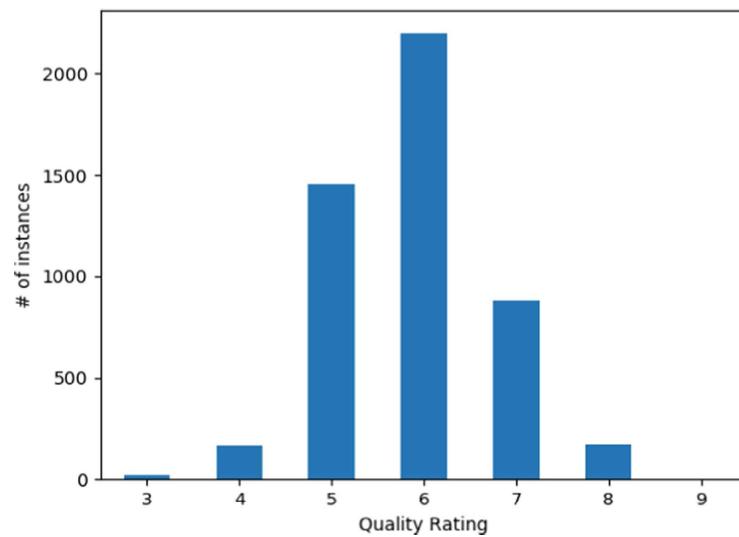
**Table 1.** Example of raw data scaling prior to normalization

Feature name	Fixed acidity	Volatile acidity	Citric acid	Residual sugar	Chlorides
Mean	6.9	0.3	0.3	6.4	0.05
SD	0.8	0.1	0.1	5.1	0.02
Max	14.2	1.1	1.7	65.8	0.35
Min	3.8	0.1	0	0.6	0.01

metric data points for each attribute were transformed by centering on the mean and scaling by the standard deviation.

### 3.2 Class Re-labeling

Initial visual exploration of the raw data shows significant asymmetry with respect to the quality class frequencies. As the following Fig. 1 shows, most wine samples have a quality rating of five or six; samples in the low and high-end ratings (3/4 and 8/9) are far less numerous and occupy the tail regions of the quality class distribution.

**Fig. 1.** Distribution of wines by quality.

In a highly competitive market, both the poorest quality and the highest quality wines would be expected to be underrepresented and rare. Wine-making businesses would have difficulty remaining financially solvent by manufacturing ‘poor’ quality wines; therefore, market selective pressure exists to always increase quality. Likewise, the highest quality wines are also the most difficult to manufacture and require the most skill and talent; therefore, these wines are expected to be rare as well. As a result, wine samples with average quality, in general, tend to be overrepresented in the marketplace and the Wine Quality dataset reflects this reality.

To overcome this challenge, the samples were regrouped into a smaller number of classes that would result in a less asymmetric distribution. Initially, ANOVA (analysis of variance) was used to determine whether the means of the attributes belonging to the various quality categories were significantly different from each other. Two attributes, (citric acid and sulphates) were identified as not significant (0.05 level) for any quality groups. All other attributes were significant for at least 5 out of the 7 quality classes. Tukey's post-hoc testing was performed for each class (i.e., 3–9). The analysis showed that the means of the two higher quality groups (7 and 8) are not significantly different from each other at the 0.05 significance level; therefore, for the current study, classes 7 and 8 were combined into one quality group representing above average/ higher quality wines. Furthermore, after performance of the outlier removal step (discussed later), it was noted that the two most extreme quality groups '3' and '9' contain only 25 and 5 observations each. Given such low representation, instances of class 3 and 9 were therefore dropped from the dataset. Table 2 shows the original quality class labels and the updated mapping to new categories.

**Table 2.** Coded class categories.

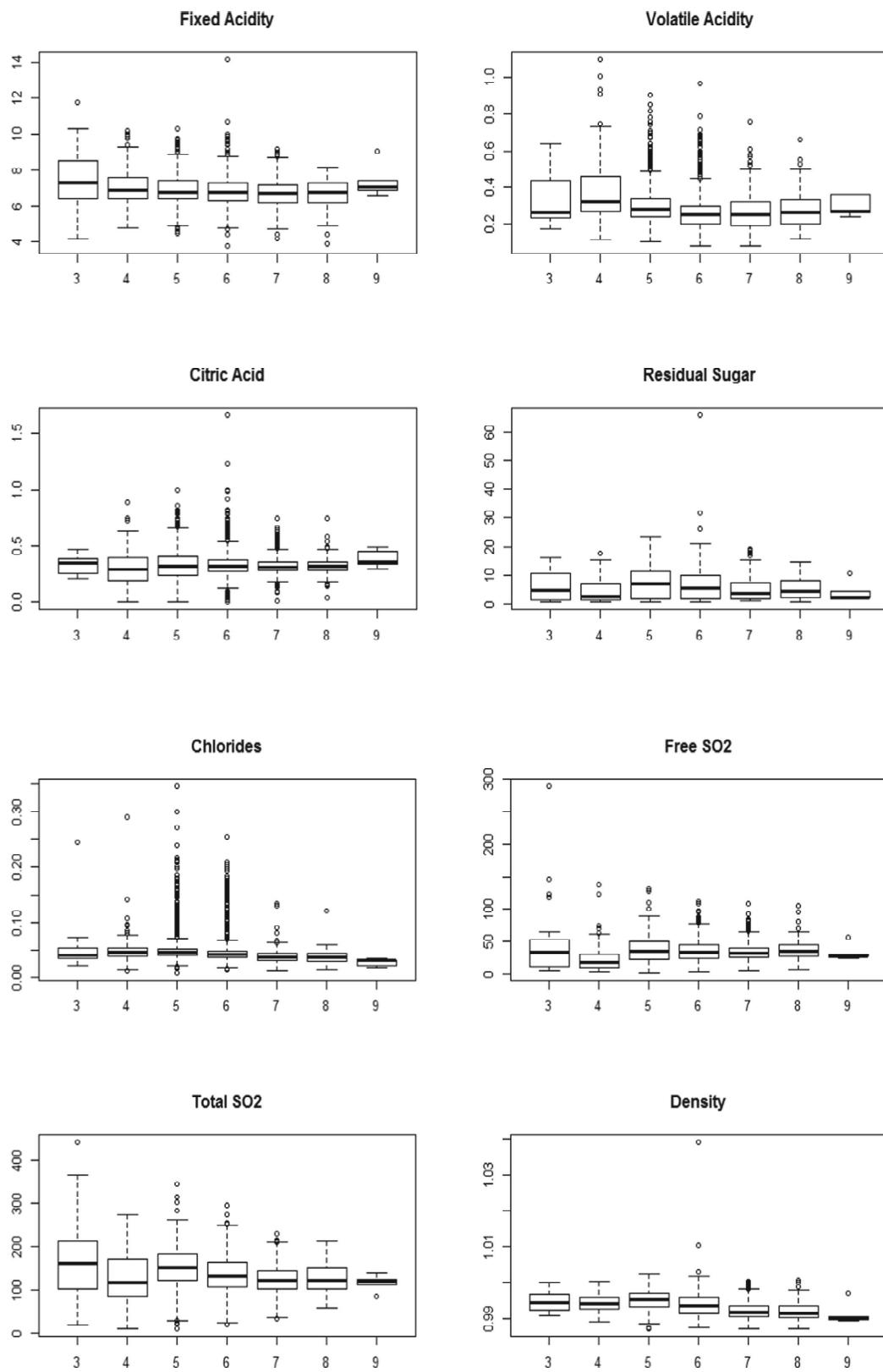
Original class	New coded class	Wine quality	N	Total (%)
4	0	Low	110	3%
5	1	Below average	1,216	28%
6	2	Average	1,984	46%
7 and 8	3	Above average/high	994	23%

As already indicated, ANOVA analysis suggested citric acid and sulphates attributes may not provide sufficient discriminatory information to contribute to overall class prediction during downstream modeling. Initial exploratory modelling (multilayer neural network) using the full feature set and the reduced feature set did not show an improvement in classification accuracy (data not shown). Therefore, the full feature set was used for all downstream modeling approaches.

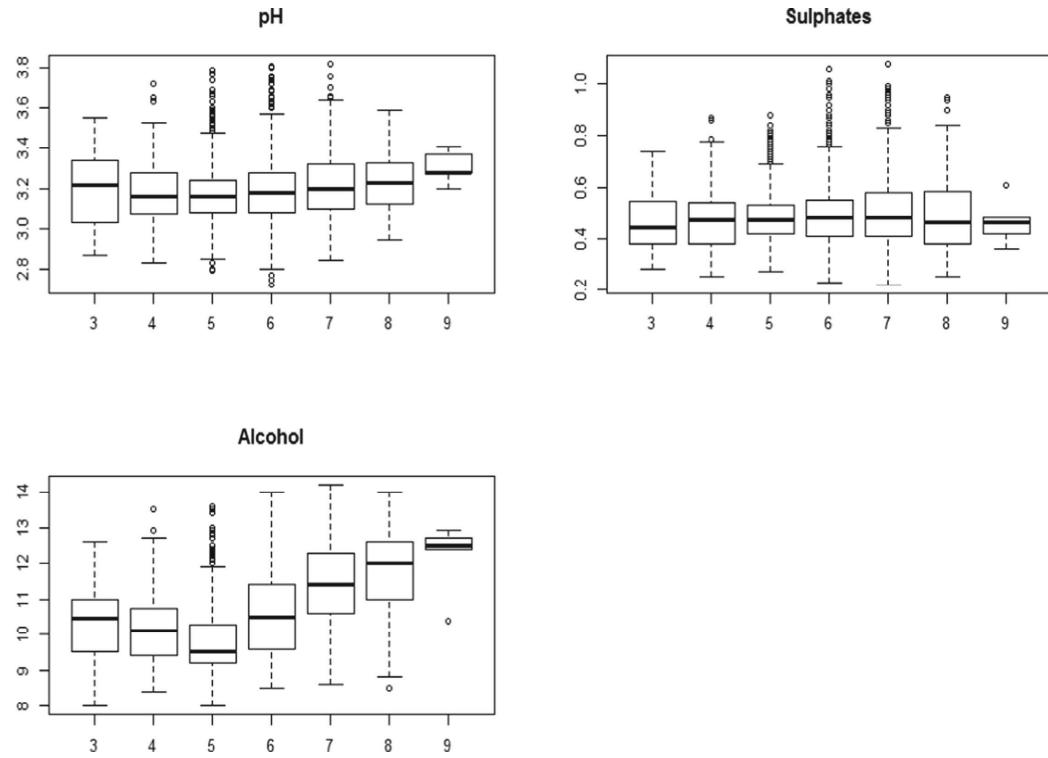
### 3.3 Outlier Detection

Boxplots were created for each attribute using the unprocessed (unscaled) raw data set in order to visually estimate the level of outliers present within each respective quality class [6, 7]. Note: Quality groups are plotted on the x-axis while the instance data for each attribute is on the y-axis. See the following Fig. 2.

As shown in Fig. 2, the majority of the eleven physicochemical attributes contain data points for one or more quality classes that are greater than 1.5 times the interquartile range (IQR) (represented as dark circles above and below whiskers). Due to space constraints, a detailed analysis of the potential outlier data was not performed; however, these data are considered to have a negative potential impact on downstream analyses if not mitigated. By convention, the 'three sigma rule' was used as support; given that the Wine Quality



**Fig. 2.** Boxplots of physicochemical attributes prior to outlier removal

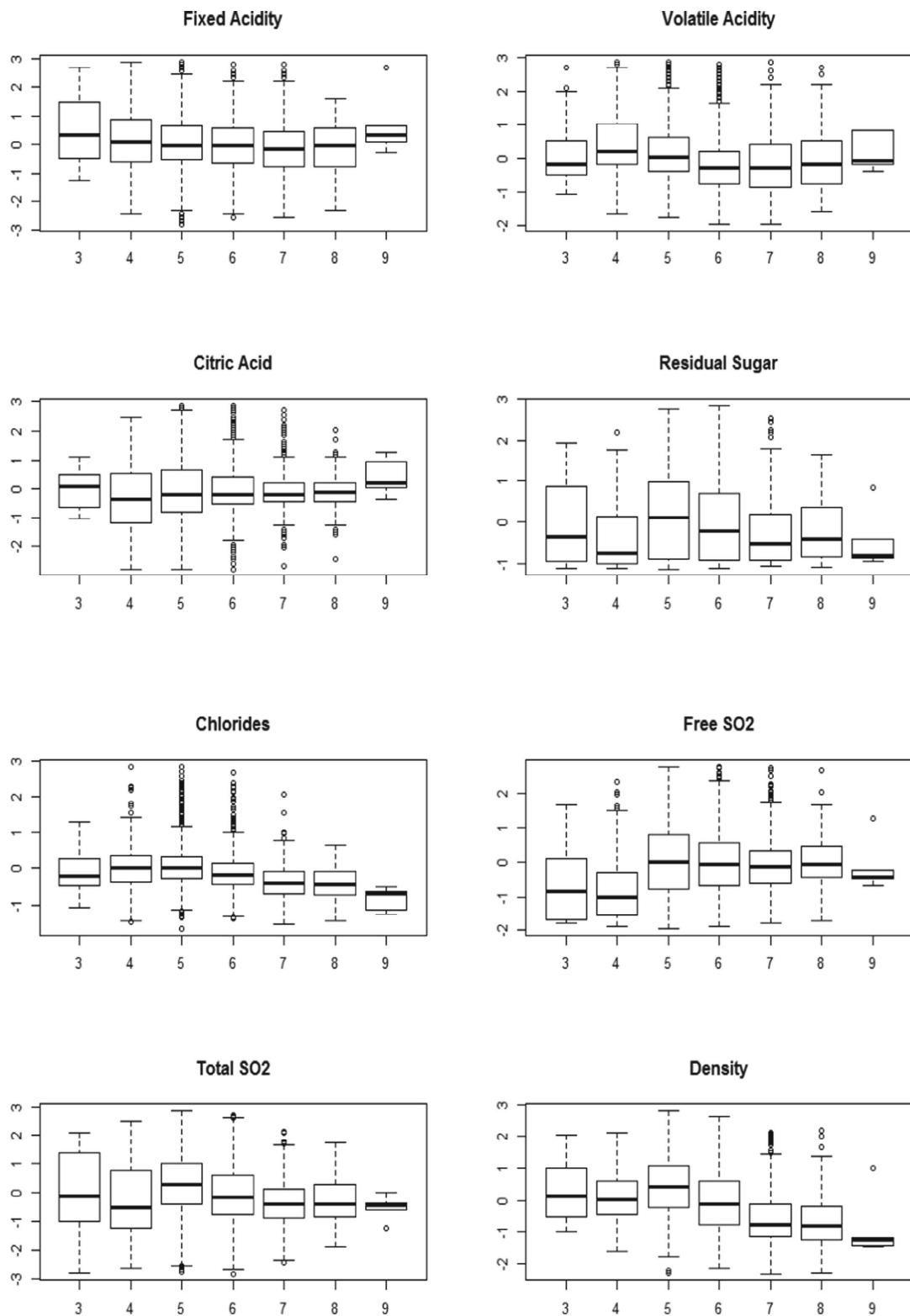
**Fig. 2.** (continued)

data is approximately normally distributed, the ratio of data points present in the set greater than  $3\text{SD}$  from the mean should be fairly close to 1:370. As a result, a normally distributed data set containing  $\sim 5,000$  instances can be expected to have approximately 14 instances  $\geq 3\text{SD}$  from the mean on average. Outlier removal was carried out on the normalized Wine Quality data set using the Spyder IDE to produce a reduced instance data set containing 4,303 observations. The original set contained 4,898 observations; therefore, 595 observations  $\geq 3\text{SD}$  from the mean were removed in total. Note that the  $\pm 3\text{SD}$  method of outlier removal is less stringent than using the  $1.5^*\text{IQR}$  method; therefore, only the most extreme values were targeted and removed.

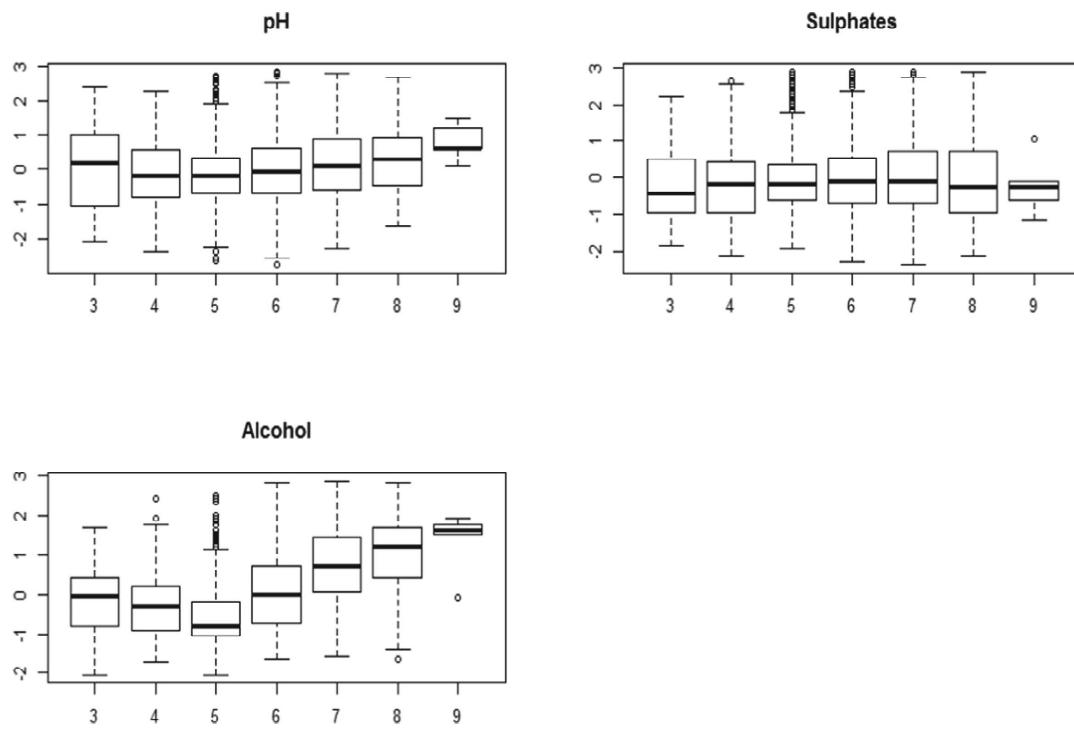
Refer to the following Fig. 3 for boxplots of attributes showing distributions after outlier removal. Note: Quality groups are on the x-axis while instance values are on the y-axis.

### 3.4 Principle Component Analysis

Principal component analysis or PCA, is a statistical procedure that uses an orthogonal transformation to convert a set of observations or attributes into a set of values of linearly uncorrelated variables termed principal components [10]. PCA is an unsupervised technique and evaluates the data set as a whole, without class labels. In general, PCA seeks to find a direction or directions that maximize the variance in the data while attempting to discover a relationship between independent variables without consideration of dependent variables.



**Fig. 3.** Boxplots of physicochemical attributes after outlier removal

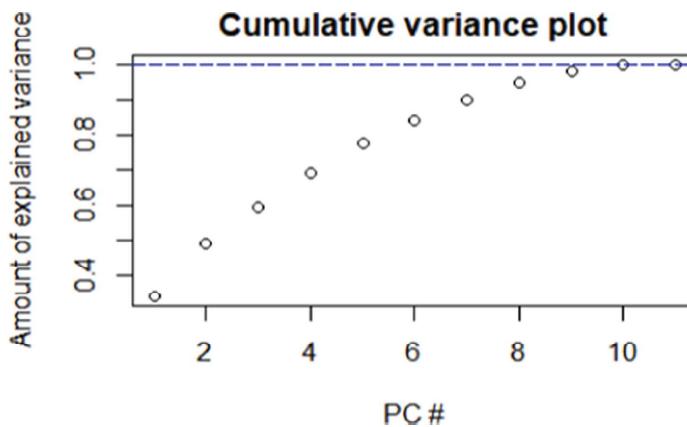
**Fig. 3.** (continued)

For the current study, PCA was performed on the normalized, reduced data set to extract the maximum variance from the complete attribute set. After PCA transformation, a set of eleven attributes (principal components) were created and used for downstream modelling. Inspection of the individual factor loadings for each component based on attribute indicates high loadings for density, total sulfur dioxide, chlorides, volatile acidity, citric acid and alcohol % (data not shown). The factor loadings explain the relationship between a given attribute and the components produced as a result of the PCA transformation and represent the percent of variance in that variable explained by the component. In general, attributes with higher factor loading are more significant to the derived component. The following Fig. 4 shows the cumulative distribution of total variance with respect to the number of PCs.

As shown in Fig. 4, components 1 through 7 account for 90% of the total variance. The remaining 10% variance is accounted for by components 8 through 11. The individual eigenvalues along with variance percent for each principle component is shown in the following Table 3.

In the following Fig. 5, the first two principle components produced from the PCA transformation are plotted showing the strength of the individual contributions for each attribute. Attributes well-separated from each other in the graph space are negatively correlated while those occupying the same space are positively correlated.

As shown in Fig. 5, alcohol%, density and residual sugar show strong factor loadings on the first principle component (Dim1). Alcohol% shows a strong negative loading while residual sugar and density show strong positive loadings on PC 1 (Dim1). As expected, density and residual sugar are positively correlated; increasing concentration of residual

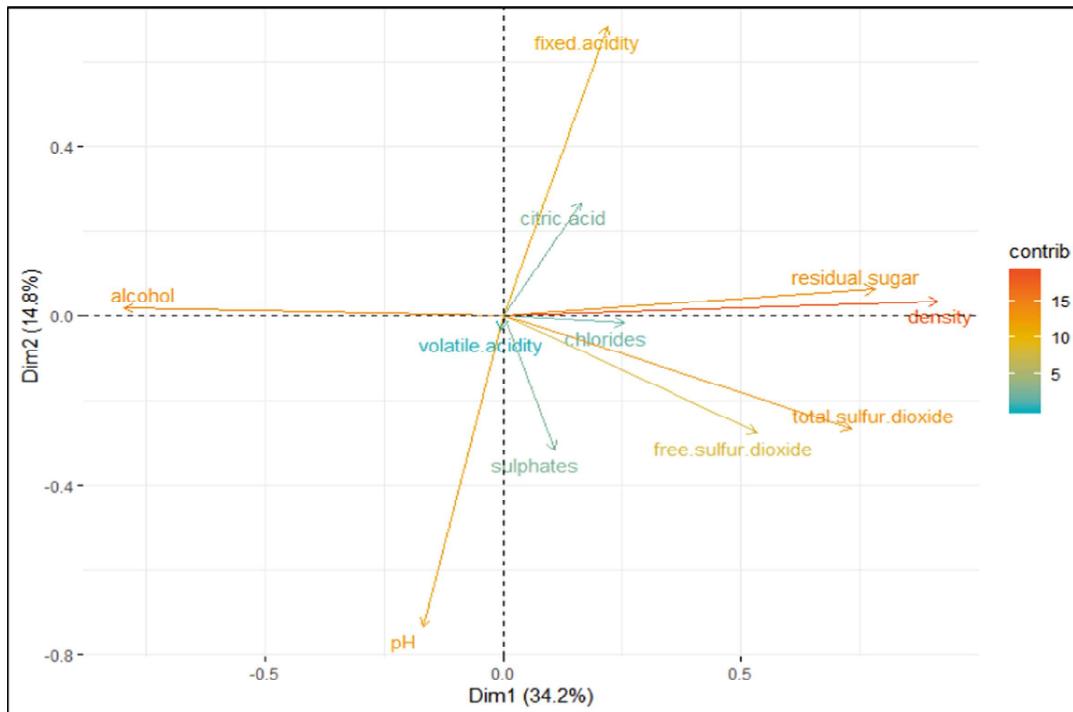
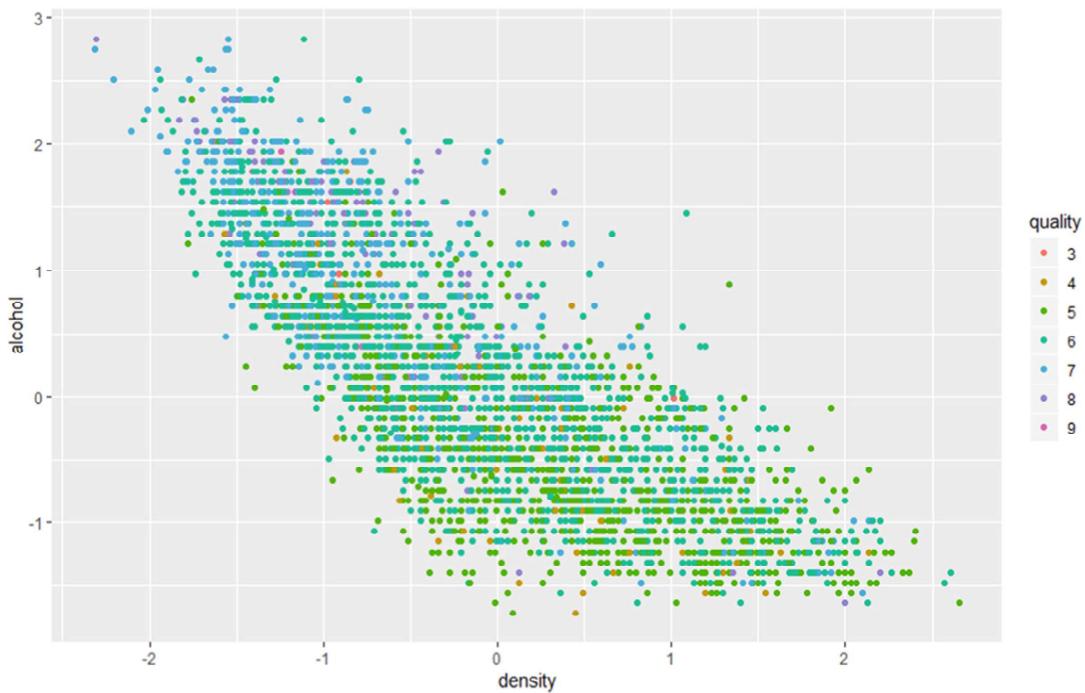
**Fig. 4.** Cumulative variance plot**Table 3.** PCA eigenvalues and variance percent for each component

Component	Eigenvalue	Variance	Cumulative variance
Dim.1	3.06	34.2%	34.2%
Dim.2	1.32	14.8%	49.0%
Dim.3	0.96	10.7%	59.7%
Dim.4	0.84	9.4%	69.1%
Dim.5	0.75	8.4%	77.5%
Dim.6	0.57	6.3%	83.8%
Dim.7	0.55	6.2%	90.0%
Dim.8	0.44	5.0%	95.0%
Dim.9	0.25	2.8%	97.8%
Dim.10	0.18	2.0%	99.9%
Dim.11	0.01	0.1%	100.0%

sugar results in increasing wine density and lower alcohol% as a result. For example, the alcohol%/ density attribute relationship is shown in the following Fig. 6.

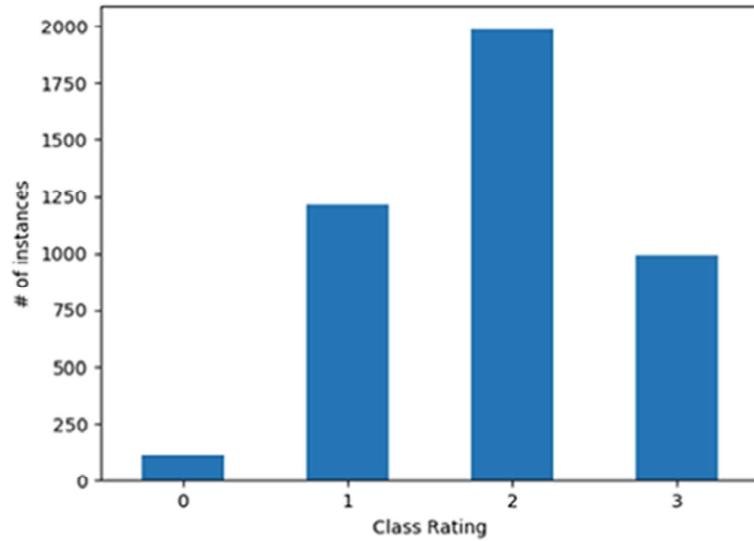
In general, incomplete fermentation results in higher density wines with higher residual sugar and lower alcohol%. As fermentation goes to completion, sugars are converted to alcohol (and CO<sub>2</sub>); as a result, the density of the wine is lower when the alcohol% is higher. As shown in previous Fig. 5, fixed acidity (positive) and pH (negative) show strong factor loadings on PC2 (Dim2). This is also expected; as fixed acidity increases, the pH will decrease. Total sulfur dioxide and free sulfur dioxide are positively correlated and demonstrate moderate factor loading on PC1. The remaining attributes show low to moderate factor loadings on PC2.

To summarize, the raw Wine Quality dataset was transformed using z-score normalization (set mean = 0, standard deviation = 1) across all attributes to place all values on the same scale. After normalization, values greater than  $\pm 3$  standard deviations from

**Fig. 5.** PCA variable plot**Fig. 6.** Scatterplot of alcohol% and density attributes

the mean for all attributes were removed from the data set. Quality groups 3 and 9 were removed due to low number of instances (~1% of dataset) and high variance. Finally,

initial ANOVA analysis performed during the data exploration phase indicated that quality groups 7 and 8 are not statistically significantly different from each other at the 0.05 significance level. Therefore, groups 7 and 8 of the original dataset were combined into one representative quality class. The dataset was then converted to a NumPy array with dimensions  $(4,303 \times 11)$ , suitable for multilayer neural network (MNN) modelling. Prior to MNN, label-encoding was applied to the quality class column. The following Fig. 7 shows a frequency histogram of samples using the new quality class categories.



**Fig. 7.** Quality classes and resulting data distributions

Note that average quality wines (class 2) represent around 50% of the transformed dataset; below average (class 1) and high-quality wines (class 3) represent 28 and 23% of the dataset, respectively. Low-quality wines (class 0) represent 3% of the dataset; this imbalance could not be adjusted in the same manner as combining quality groups 7/8 since initial ANOVA analysis of group means showed that class 0 is significantly different from all other classes at alpha 0.05 (data not shown). An oversampling approach could have been used at this point to synthetically create additional instances to further adjust the class imbalance. Given the level of preprocessing already performed on the data set, further processing was therefore deemed unwarranted for the current study. One hot encoding was performed on the new class labels, which is required by MNN models. Prior to training the downstream classifiers, label-encoded and one-hot encoded datasets were split into 80/20 training and testing datasets, respectively.

## 4 Modeling Systems

The primary goal of the current study is the evaluation of various modeling systems for the prediction of wine quality based on an ordered categorical attribute representing four discrete quality levels, 0 through 3. Therefore, the output for all models tested in this study is a classification of 0 (low), 1 (below average), 2 (average) or 3 (above

average/ high). The specific model systems used in the current study are briefly outlined as follows:

1. Support Vector Machine

A Support Vector Machine (SVM) is a discriminative classifier that creates a separating hyperplane between two or more classes in the feature space. The hyperplane is then used as the basis for separation of new class instances. SVM seeks to find an optimal hyperplane that maximizes the margin between classes.

2. Decision Tree

Decision Trees (DTs) are hierarchical models for supervised learning, where the local region is identified in a sequence of recursive splits in a smaller number of steps. It is built up of decision nodes where each node implements a test function with an outcome. The process begins at the root and is repeated until a leaf node is reached; the value within the leaf is the output [8].

3. Multilayer Neural Network

A Multilayer Neural Network (MNN) is a feed-forward artificial neural network that contains several layers of artificial neurons or nodes. The MNN will generally have one input layer, one hidden layer and one output layer; however, different numbers of layers are possible depending on the application. The MNN can distinguish data that is not linearly separable and employs a variety of different optimizers and hyperparameter tuning options.

4. Random Forest

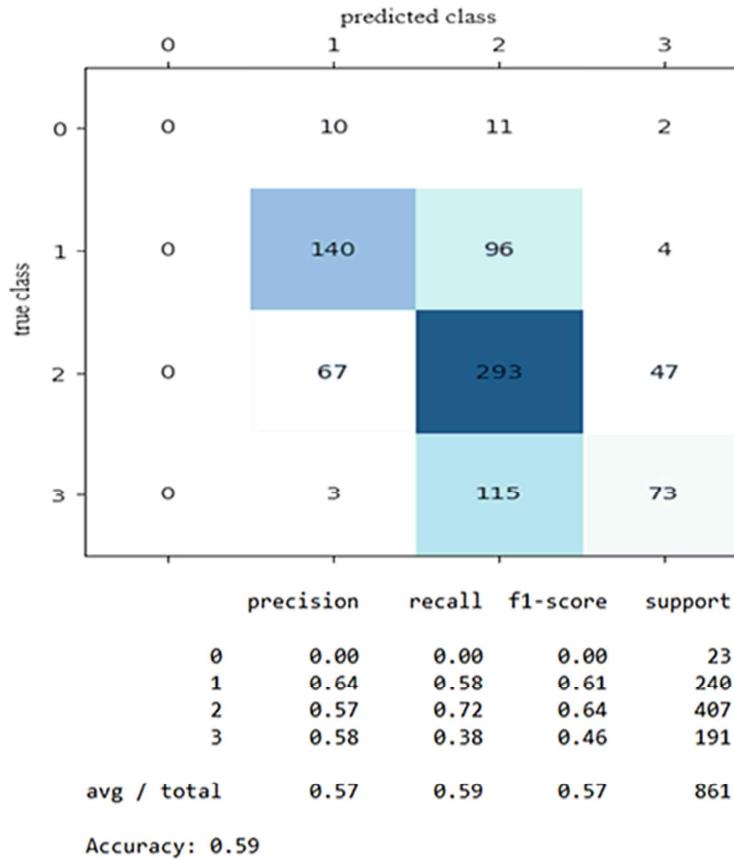
Random Forests (RFs) are an ensemble learning method used for classification and regression. The algorithm functions by constructing a predetermined number of decision trees during the training phase and outputs the class that represents either the mode of the classes (classification) or mean prediction (regression) of the individual trees. More specifically, Random Forests are a collection of Decision Trees where each tree is slightly different from the others due to the random nature of the attributes chosen to construct an individual tree. In general, each tree can be a good predictor but may have a tendency to overfit on part of the data. Building many decision trees based on randomly sampled training data and then averaging the results can avoid overfitting yet result in a model with higher classification accuracy [8, 9].

The following sub-sections present details of applying these methods on the processed white wine data taken from the Wine Quality dataset.

#### **4.1 SVM (Support Vector Machine)**

The SVM module provided in the scikit-learn library is used to instantiate an SVM classifier object. Three different kernel types (polynomial, Radial Basis Function (RBF) and sigmoid) were tested, resulting accuracies compared, and best results presented. Note that kernels represent mathematical functions used to transform the input data and provide an evaluation based on ‘similarity’. Among those tested, the RBF kernel achieved the highest accuracy of 59%. The confusion matrix, along with standard metrics (i.e., precision, recall, F1-score) are shown in the following Fig. 8.

Note that rows in Fig. 8 represent the true sample class for each of the four classes, whereas, columns represent predicted classes by the model for each of the four classes.

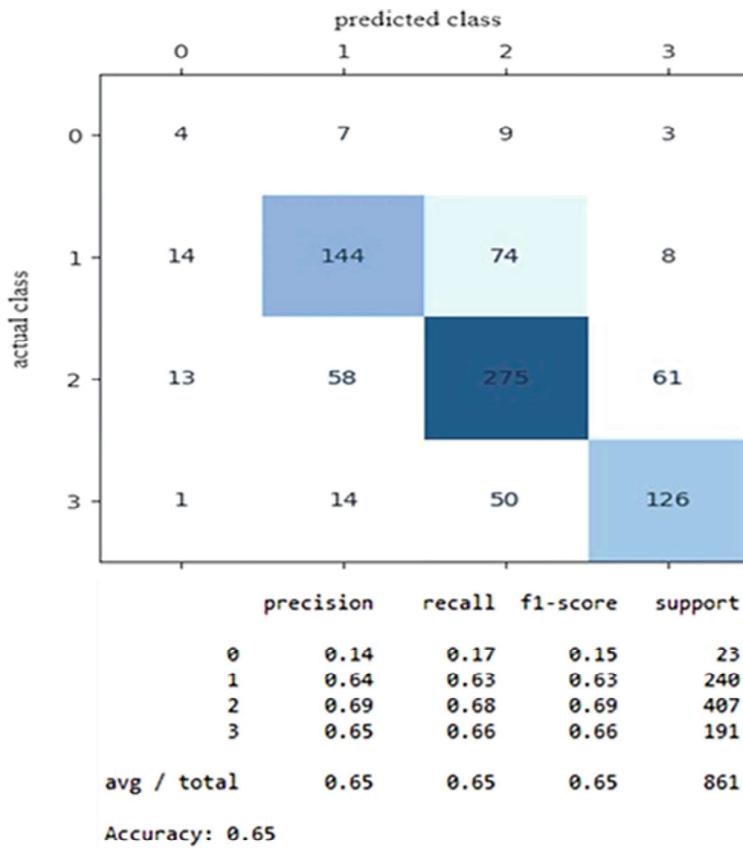


**Fig. 8.** SVM classification results

For example, the value of zero in cell (0, 0) in Fig. 8 shows that none of the true class zero samples were successfully classified as class zero. For the SVM model, all of the true class zero samples are misclassified, 10 as class-1, 11 as class-2, and 2 as class-3. Diagonal cells represent the number of samples that have been correctly classified, for example, 293 samples of true class-2, have been correctly classified by the SVM model. Note that the color intensity provided in each matrix panel is directly related to the number of instances correctly classified. Higher numbers of correctly classified instances correlate with darker coloration inside the respective panel. A perfect classification (i.e., 100%) would show dark colored panels along the diagonal from top left to bottom right; the remaining panels would be white and filled with zeros. Panels outside of the main diagonal represent misclassified instances and follow the same color scheme.

## 4.2 Decision Tree

The scikit-learn decision tree package is used to instantiate a decision tree object, with a max depth of 15 and the entropy criterion used to determine the split at each node. The decision tree model achieved a classification accuracy of 65%, which is 6% higher than the SVM model discussed previously. The confusion matrix, along with standard metrics are shown in the following Fig. 9.



**Fig. 9.** Decision tree classification results

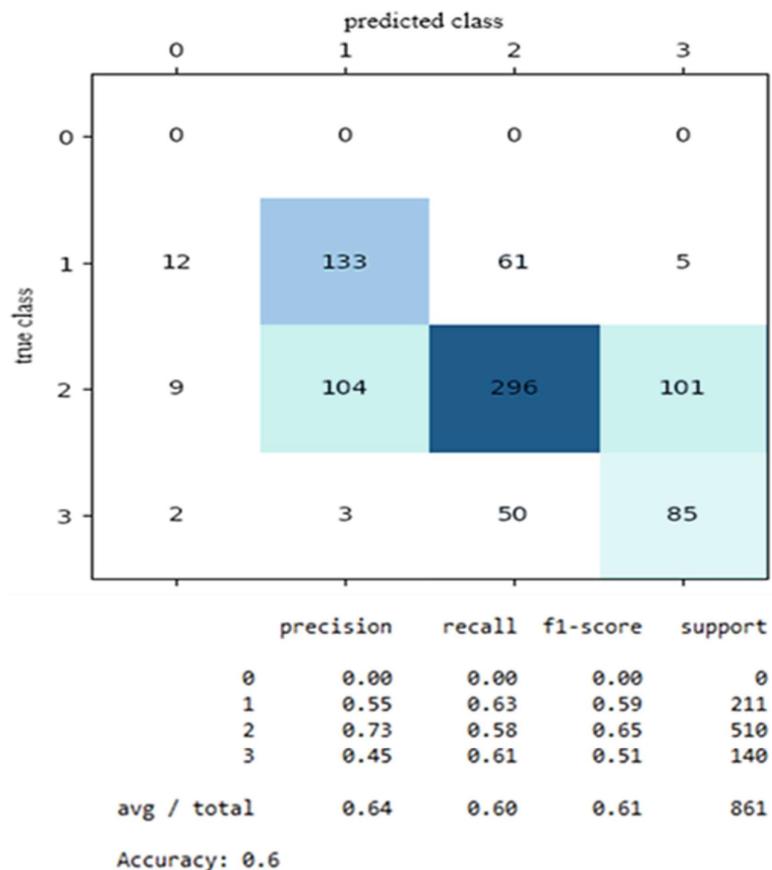
### 4.3 MNN (Multilayer Neural Network)

An MNN employing stochastic logistic regression is performed on the final processed data set using Keras [11]. Keras is a high-level neural networks API (Application Programming Interface), written using the python programming language. The following parameters are used for the first MNN model:

1. One dense input layer, input shape = 11, output = 1,000 with ‘ReLU’ activation function.
2. One dense hidden layer, output = 500 with ‘ReLU’.
3. One dense hidden layer, output = 250 with ‘ReLU’.
4. One dense output layer, output = 4 with ‘Softmax’ activation and L2 regularization of 0.0001.

During model compilation, stochastic gradient descent (SGD) was used as the optimizer. The model was trained for 75 epochs and resulted in a 60% classification accuracy. Some additional hyper-parameter trial and error tuning (not shown) did not result in significant improvement in the classification accuracy. The confusion matrix along with standard metrics are shown in the following Fig. 10.

Given these results, a second MNN classification was performed employing stochastic logistic regression with the same parameters as in the first MNN model; however,



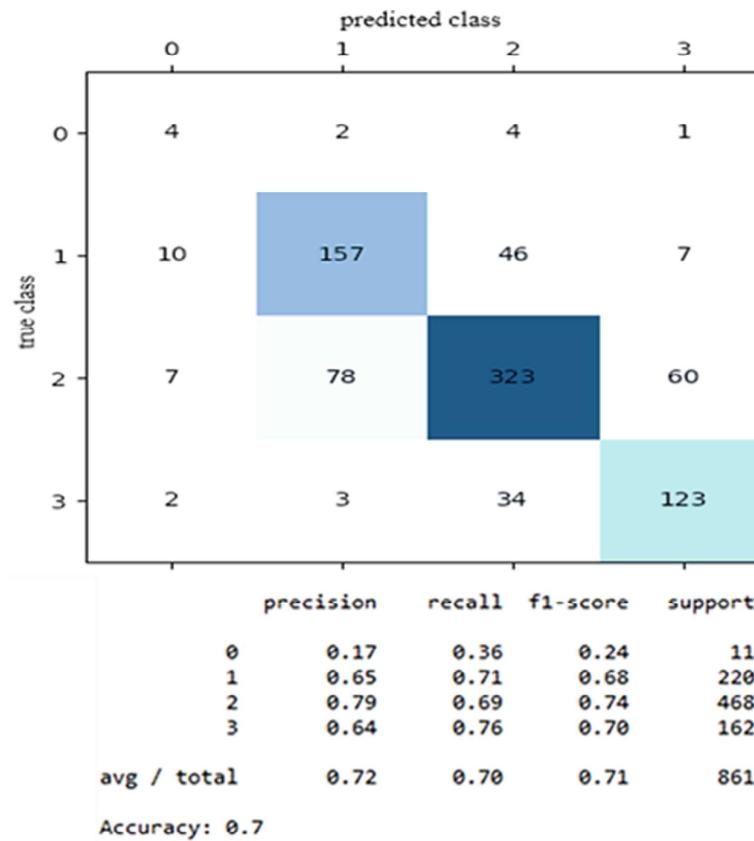
**Fig. 10.** Multilayer neural network classification results (SGD)

the ADAM (Adaptive Moment Estimation) optimizer was used during model compilation instead of SGD (stochastic gradient descent). The second model was trained for 75 epochs and resulted in an improved 70% classification accuracy over the previous MNN model. The confusion matrix, along with standard metrics are shown in the following Fig. 11.

#### 4.4 RF (Random Forest Ensemble)

Similar to the other modeling techniques, a scikit-learn Random Forest Classifier was used, and a RandomForestClassifier object instantiated with 21 estimators using entropy as the criterion value. Entropy measures the quality of a potential split and determines the information gain at a potential splitting point; tree nodes with higher information gain will be selected for splitting. Information gain and entropy are inversely related as higher information gain results from a lower entropy split thereby resulting in a more homogeneous branching. The random state was selected at a value of 90 for this classifier and is used to ensure the uniformity of the train/test splitting for the indices. Setting the random state to a value greater than 0 ensures reproducibility of results.

The random forest model achieved a classification accuracy of 72%, which represents a 2 percent improvement over the best accuracy achieved by the MNN with the ADAM



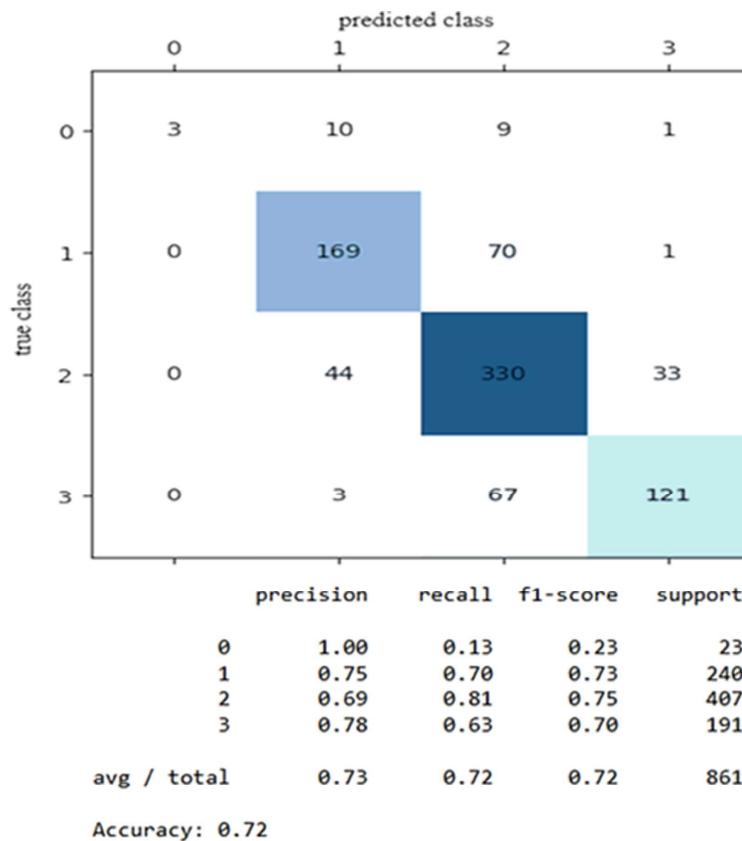
**Fig. 11.** Multilayer neural network classification results (ADAM)

optimizer. The confusion matrix, along with standard metrics are shown in the following Fig. 12.

## 5 Results

Table 4 summarizes and compares prediction accuracy and other metrics for all modeling techniques. As can be observed, the Random Forest classifier achieved a prediction accuracy of 72%, which is the highest accuracy among all models used.

Note that average precision, recall and f1-score for each model tested are also shown in the previous Table 4. The Random Forest classifier also achieved the highest average precision and recall of all models tested. The next best classifier was the Multilayer Neural Network using the ADAM optimizer. The MNN model achieved this accuracy after 75 epochs; given the options available in Keras, further hyper-parameter tuning and experimentation is possible. Increasing the number of hidden layers could also be tested. Both SVM and the MNN using stochastic gradient descent as the optimizer scored the lowest accuracies on the white wine dataset (59% and 60%, respectively). As indicated, further hyperparameter tuning could be applied to this model, or other optimizers could be chosen. Results from other studies using various classification as well as regression models applied to the white wine dataset are summarized and presented in Table 5. Best results are highlighted in bold.

**Fig. 12.** Random forest results**Table 4.** Model performance comparison

Model	Parameter	Class. accuracy	Avg. precision	Avg. recall	Avg. F1 score
SVM	RBF kernel	59%	0.57	0.59	0.57
MNN	SGD optimizer	60%	0.64	0.60	0.61
Decision tree	Max. Depth = 15	65%	0.65	0.65	0.65
MNN	ADAM optimizer	70%	0.72	0.70	0.71
Random forest	Entropy, est. = 21	72%	<b>0.73</b>	<b>0.72</b>	<b>0.72</b>

Average classification accuracies reported by Cortez et al. are highest for the SVM regression model at 64.6%, while MNN and MLR model results are not significantly better than a random classification. Accuracies reported by Appalasamy et al. using both decision tree (ID3) and Naive Bayes classifiers are 52.3 and 50.5%, respectively. In contrast, results achieved in the current study are 70% for MLP/ADAM and 72.4% for RF which represent a significant improvement as compared with these results.

**Table 5.** Model performance comparison

Study	Pre-Processing	Model	Task	Avg. accuracy
P. Cortez et al. [2]	FS <sup>2</sup>	SVM	Regression	<b>64.6%</b>
		MNN	Regression	52.6%
		MLR	Regression	51.7%
P. Cortez et al. [12]	FS <sup>2</sup>	SVM	Regression	<b>64.3%</b>
		MNN	Regression	52.6%
		MLR	Regression	51.7%
Appalasamy et al. [3]	FS <sup>2</sup>	DT (ID3)	Classification	<b>52.3%</b>
		Naïve Bayes	Classification	50.5%
Er et al. [14]	PCA <sup>4</sup>	RF	Classification	<b>70.3%</b>
		RF	Classification	68.7%
Hu et al. [13]	CR <sup>1</sup> , FS <sup>2</sup> , SMOTE <sup>5</sup>	SVM	Classification	<b>93.7%</b>
		DT	Classification	92.0%
Aich et al. [16]	FS <sup>2</sup>	Multiple models	Classification	<b>98.8%</b>
Aich et al. [17]	FS <sup>2</sup>	Multiple models	Classification	<b>99.6%</b>
Gupta [15]	FS <sup>2</sup>	MNN	Regression	NR <sup>6</sup>
		SVM	Regression	NR <sup>6</sup>
Current study	CR <sup>1</sup> , OR <sup>3</sup> , PCA <sup>4</sup>	RF	Classification	<b>72.4%</b>
		MNN (Adam)	Classification	70.0%
		DT	Classification	64.2%
		MNN (SGD)	Classification	60.0%
		SVM	Classification	58.7%

<sup>1</sup> CR: Class reduction<sup>2</sup> FS: Feature selection<sup>3</sup> OR: Outlier removal<sup>4</sup> PCA: Principle components analysis<sup>5</sup> SMOTE: Synthetic minority oversampling<sup>6</sup> NR: Accuracy, precision, recall and F1-score not reported

MLP/ADAM represents a 5.4% classification accuracy improvement and RF represents a 7.8% improvement over the best achieved result of 64.6%, as reported by Cortez et al. using the SVM regression-based model. Er et al., reported 70.3% accuracy using a RF model which compares with the current study accuracy of 72.4% using RF also.

Hu et al., reported a best accuracy of 93.7% using an SVM model in combination with SMOTE pre-processing prior to classification. As mentioned earlier in this paper, SMOTE is an oversampling technique that generates synthetic samples from the minority class [18]. SMOTE is used to obtain a synthetically class-balanced training set, which is then used to train the downstream classifier. SMOTE can be used in combination with

majority class undersampling to remedy asymmetric class distributions. The SMOTE technique is not without issue however, as it will generalize the minority area without reference to the majority class. The sampling space is confined to the area nearest the minority class instances. If the majority and minority classes are not well separated in the feature space, SMOTE could potentially produce synthetic samples representative of the majority class, but labeled as the minority class (i.e., ‘class mixture’). For high dimensional data, this issue could be especially problematic. In this regard, the application of SMOTE can therefore result in overly optimistic accuracies. Additionally, since SMOTE synthesizes minority samples similar, but not identical to the original data, the technique can create additional noise that may compound with noise already present within the data set. For the Wine Quality data set, it is known that most attributes contain high variance instances that exceed three standard deviations from the mean, i.e., noisy data. As shown in the current study, using z-score normalization, 12% of instances, or a total of 595 observations exceeded this measure and were therefore discarded prior to downstream modeling.

Similar to Hu et al., Aich et al., also report high classification accuracies (99.6%) using multiple classification models applied to the Wine Quality dataset in two separate studies. It should be noted however, that the authors elected not to report the confusion matrices, feature selection analysis details or exact attributes comprising the reduced feature sets. Additionally, no pre-processing (scaling, normalization, outlier detection, PCA, etc.) steps are indicated in either [16] or [17]; however, as shown in the current study, the Wine Quality data set demonstrates significant class imbalance and a high disparity with respect to individual attribute scale. These issues are known to affect classification accuracy but were not reported in either [16] or [17].

Gupta reports low training, testing and validation errors for MLP and SVM regression analysis, but does not provide the corresponding confusion matrices, or precision/recall/ F1 score metrics for consideration. Similar to Aich et al., Gupta performs feature selection using a regression approach and indicates that pre-processing is recommended for the Wine Quality data, yet does not report whether or not preprocessing (other than feature selection) was performed.

## 6 Discussion

As compared to several recent works where classification metrics are reported, the current study results are therefore considered successful in that higher classification accuracies, precision and recall metrics were obtained using the random forest model. To the best of our knowledge, the current study is the first to implement outlier mitigation and quality class reduction prior to downstream modeling.

Other studies report high classification accuracies (>95%); however, three studies do not provide classification metrics to enable a clear comparison and one study uses synthetic data oversampling (SMOTE) to boost the lower quality classes prior to analysis. In general, oversampling the minority class by creating synthetic instances can result in overly optimistic accuracies. For the current study, we preferred to use the available data and accept the lowest quality class imbalance.

Interestingly, the SVM model used in the current study achieved a lower average classification accuracy of 58.7% compared to 64.6% achieved by Cortez et al. The

difference in accuracy may be attributed to the fact that the SVM model used by Cortez et al. is applied to the data set as a regression task where continuous output was expected. For example, 3.9, 4.0, and 4.1 output values could all be interpreted by the regression model as class ‘4’ but not class ‘5’ or ‘6’, etc. In contrast, the current study used the SVM to solve a discrete classification task, where Class 1 = 0, Class 2 = 1, etc. Therefore, the difference in the two approaches may have contributed to the difference in average classification accuracies.

A review of the study design implemented by Cortez et al. shows that the authors used all seven quality groups as input to the SVM regression model (3 through 9), whereas the current study condensed the quality groups into four representative classes (0 through 3) in an attempt at class size reduction and overall model parsimony. Similar to the current study, Hu et al., condensed the quality groups into three classes, low (3/4), normal (5/6/7) and high (8/9). By aligning and condensing the quality groups in this way, Hu et al., create a greater class imbalance since quality group 7 is included in the ‘normal’ or mid-range group. As related to the original scale (0–10), group 7 is above average, along with group 8 which is even higher. As mentioned earlier, initial ANOVA analysis showed that the group means for quality classes 7 and 8 were not significantly different at an alpha of 0.05. The analysis therefore supports a design where groups 7 and 8 are simply combined into one group representing ‘higher’ quality wines including wines that are ‘above average’ thereby alleviating the class asymmetry to some degree. Given the design implemented by Hu et al., SMOTE oversampling would therefore be required in order to relieve the quality class asymmetry and enhance the downstream predictive classification.

To the best of our knowledge, one issue not addressed by any related works is the large range of the original rating scale (0–10) used to score the wine samples. The current study showed that condensing the quality groups most likely boosted the classification accuracies. After in depth data set exploration and analysis, the current study suggests that the eleven physicochemical attributes comprising the Wine Quality data set do not contain sufficient discriminatory power to allow a high level of differentiation between eleven distinct quality groups. Issues such as class asymmetry, high variance instances and low statistical power negatively affected the classification accuracy in the current study even with pre-processing performed. Quality scale revision and reduction (maximum of five ratings for human sensory evaluations, i.e., low, below average, average, above average and high) could therefore be one approach to address the data set limitations in the future. Additional measureable wine attributes such as malic acid content, magnesium level, phenol content, flavonoid content, color intensity and hue could be incorporated along with the standard physicochemical tests used in the current study to supplement the data and potentially provide increased quality class differentiation. We refer the interested reader to the following dataset [19].

## 7 Conclusion

For the current study, one variant (white) of the Portuguese “vinho verde” wine was analyzed and comprised a data set with 4,898 instances and eleven physicochemical attributes (i.e., features). During the data exploration phase, a number of pre-processing

steps were performed (i.e., outlier detection and removal, PCA, class size reduction, etc.) in order to reduce the variance between classes for each attribute and thereby achieve greater class separation. The overall intent was to improve the classification accuracy for the models tested. After pre-processing steps were completed, four different classifiers were implemented for quality prediction; out of these, the Random Forest and MNN (with ADAM optimizer) classifiers outperformed the SVM and Decision Tree models. Implementation of each algorithm was carried out using the Python scikit-learn library.

The results obtained in the current study represent an improvement in classification accuracy, precision and recall, as compared to four previous related studies that also provided similar evaluation metrics [2, 3, 12, 14]. Other related studies report very high accuracies that exceeded the current study results ( $>95\%$ ), but unfortunately do not report the standard classification metrics necessary for straightforward comparison [15–17]. Another study [13] reports high classification accuracy (93.7%); however, the authors used the SMOTE oversampling technique to boost the extreme classes by creating synthetic data. After some consideration, in lieu of obtaining additional measured data for the extreme classes, the SMOTE technique could represent a very efficient method for mediating significant class asymmetry in real world data sets. We would therefore like to see additional work done to provide more validation for the approach.

The Random Forest model used in the current study resulted in the highest classification accuracy of all models tested for the white wine dataset (i.e., 72.4%). With further training on new data, the accuracy of the Random Forest classifier is predicted to increase; therefore, this algorithm could be used in the wine industry for predicting the quality of white (and red) wines during the initial evaluation phase. The basic physicochemical tests (i.e. density, pH, alcohol %, etc.) used to create the dataset are generally applicable to all wines, regardless of type or manufacturer, and can be performed by a standard testing laboratory. The current study results indicate that objective assessments of wine quality using measurable physicochemical attributes modeled using machine learning techniques can align with and support sensory-based human expert evaluations.

## 8 Future Work

Given the high asymmetry of the Wine Quality dataset with respect to the number of samples in each quality class, and the high variance of some instances within each feature set, additional work is therefore needed by the community to build a more uniformly distributed dataset with equal representation over the entire quality spectrum, i.e. poor to excellent. In particular, more representation is needed for those wines that exist at the extremes of the quality spectrum; however, given that wine quality suggests a normal distribution, acquiring data for the extremes may prove challenging. Based on the results of the current study, it is thought that a more balanced dataset could significantly improve future prediction accuracies by eliminating class bias, and allow for increased standardization in wine scoring.

Additional experiments involving trial and error hyper-parameter tuning of the MNN model along with the use of other optimizers such as *RMSprop*, *Adagrad*, *Adamax*, *Nadam*, etc. are planned for future studies. Systematic and automated hyper-parameter tuning techniques will be utilized to obtain better accuracy results for the MNN and

Random Forest modeling techniques. Aside from data set pre-processing techniques (i.e. feature selection, PCA, outlier remediation, SMOTE oversampling, etc.), hyperparameter tuning provides significant opportunity for improvement. To meet the study time constraints however, minimal hyperparameter tuning was performed during testing and only two optimizers were used.

To the best of our knowledge, other issues not addressed by the current study or related works include high multicollinearity between certain variables (i.e., residual sugar and density for example) as well as high levels of negative correlation between the strongest predictors (i.e. alcohol% and density). Features displaying high multicollinearity can negatively affect downstream modeling. Removal of one of the attributes is therefore recommended in the event that two variables explain the same component. For example, the residual sugar attribute data could be dropped since density provides the same information value. More in depth analyses to address these types of issues during the pre-processing phase is planned for future work. In particular, experiments where collections of different reduced feature sets are produced and used in downstream multilayer neural network modeling are suggested.

Feature selection methods conducted by several related studies applied to the Wine Quality dataset show some agreement related to determination of the most important input features. Cortez et al., [12], report alcohol%, sulphates, pH, free sulfur dioxide and volatile acidity as the top five input attributes, Hu et al., [13] report volatile acidity, free sulfur dioxide, alcohol%, citric acid and pH, Gupta [15] reports volatile acidity, chlorides, free sulfur dioxide, total sulfur dioxide and pH, while Appalasamy et al., [3] conclude that alcohol% and volatile acidity are the two most important predictors. Additional future work is therefore planned to experiment with different feature selection methods as applied to the Wine Quality data set and other related data sets. The related works show that different feature selection methods produce slightly different results and therefore different feature input lists. As a final note, for the current study, during the initial data exploration phase, feature selection was performed using RFE (recursive feature elimination) and RF (random forest) approaches. Initial downstream modeling attempts using a few reduced feature sets did not improve the classification metrics as compared with the full feature set (data not shown). More work is therefore planned in the area of feature extraction in order to potentially standardize the approaches and increase the accuracies.

## References

1. Hopfer, H., Nelson, J., Ebeler, S., Heymann, H.: Correlating wine quality indicators to chemical and sensory measurements. *Molecules* **20** 8453–8483 (2015). <https://www.mdpi.com/1420-3049/20/5/8453/pdf>. Accessed 8 Oct 2018
2. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J.: Modeling wine preferences by data mining from physicochemical properties. *Decis. Support Syst.* **47**, 553–574 (2009). <https://www.stat.auckland.ac.nz/~stats330/datasets.dir/cortez2009.pdf>. Accessed 8 Oct 2018
3. Appalasamy, P., Mustapha, A., Rizal, N.D., Johari, F., Mansor, A.F.: Classification-based data mining approach for quality control in wine production. *J. Appl. Sci.* **12**, 598–601 (2012). <https://scialert.net/abstract/?doi=jas.2012.598.601>. Accessed 8 Oct 2018
4. Ebeler, S.: Linking flavour chemistry to sensory analysis of wine. In: *Flavor Chemistry — Thirty Years of Progress*, Kluwer Academic Publishers, pp. 409–422 (1999)

5. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/Wine+Quality>. Accessed 12 Oct 2018
6. Muller, A.C., Guido, S.: Introduction to Machine Learning with Python, pp. 85–88. O'Reilly Media, Sebastopol (2017)
7. Anaconda Software Distribution: Computer software. Vers. 2-2.4.0. Anaconda, November 2016. <https://anaconda.com>. Accessed 12 Oct 2018
8. Alpaydin, E.: Introduction to Machine Learning. Massachusetts Institute of Technology, Cambridge (2014)
9. Raschka, S.: Python Machine Learning, pp. 128–138. Packt Publishing, Birmingham (2015)
10. Lever, M., Krzywinski, N.A.: Principal component analysis. *Nat. Methods* **14**, 641–642 (2017). State Statistical Software (2010). [Computer Software]., Nature Publishing Group, a division of Macmillan Publishers Limited. June 28th
11. Keras. <https://keras.io/>. Accessed 12 Oct 2018
12. Cortez, P., Teixeira, J., Cerdeira, A., Almeida, F., Matos, T., Reis, J.: Using data mining for wine quality assessment. In: Gama, J., Costa, V.S., Jorge, A.M., Brazdil, P.B. (eds.) Discovery Science. DS 2009. LNCS, vol. 5808. Springer, Heidelberg (2009)
13. Hu, G., Xi, T., Mohammed, F., Miao, H.: Classification of wine quality with imbalanced data. In: IEEE International Conference on Industrial Technology (ICIT), pp. 1217–1712 (2016)
14. Er, Y., Atasoy, A.: The classification of white wine and red wine according to their physico-chemical qualities. *Int. J. Intell. Syst. Appl. Eng.* 23–23 (2016). <https://doi.org/10.18201/iji.sae.265954>
15. Gupta, Y.: Selection of important features and predicting wine quality using machine learning techniques. *Procedia Comput. Sci.* **125**(2018), 305–312 (2017)
16. Aich, S., Al-Absi, A., Hui, K., Sain, M.: Prediction of quality for different type of wine based on different feature sets using supervisor machine learning techniques. *Trans. Adv. Commun. Technol. (TACT)* **7**(3), 1122–1127 (2018). <https://doi.org/10.23919/ictact2019.8702017>
17. Aich, S., Sain, M., Yoon, J.-H.: Prediction of Different Types of Wine Using Nonlinear and Probabilistic Classifiers. In: Krishna, A.N., Srikanthaiah, K.C., Naveena, C. (eds.) Integrated Intelligent Computing, Communication and Security. SCI, vol. 771, pp. 11–19. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-10-8797-4\\_2](https://doi.org/10.1007/978-981-10-8797-4_2)
18. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)
19. UCI. <https://archive.ics.uci.edu/ml/datasets/Wine>. Accessed 13 Jan 2018