IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 26, Issue 3, Ser. 4 (May. – June. 2024), PP 01-08
www.iosrjournals.org

"Predicting Heart Failure Circumstances: A Machine Learning Approach For Improved Diagnosis"

Adesina-Adebayo Fatimat Opeyemi[1] Oyewole Bernice Yetunde[2]
[1](Department Of Computer Studies, Faculty Of Science/ The Polytechnic Ibadan, Nigeria
(Federal University Oye Ekiti.)

Abstract:
This work uses a dataset of 299 patient instances with 13 parameters from the UCI machine learning library to predict the optimal model and classifier for analyzing heart failure circumstances. Explanatory Data Analysis (EDA), which uses pre-processing, spread and location metrics, and visualization approaches to make sense of the data, is where the study starts. Using the original, unbalanced, and significant feature datasets, machine learning methods including Naïve Bayes, Logistic Regression, Support Vector Machine, Random Forest, K-nearest neighbor, and Neural Network are used to determine which classifier and model works best. Imbalanced data is addressed by sampling techniques like SMOTE and ADASYN, while the most significant qualities are identified via feature selection approaches like Chi-Square, Mann Whitney U, and Random Forest.

Key Word: *Imbalanced data; Mann Whitney U; SMOTE; ADASYN; Chi-square; Random Forest.*

Date Of Submission: 03-06-2024 Date Of Acceptance: 13-06-2024

I. Introduction

A dangerous medical disease known as heart failure occurs when the heart is unable to pump enough blood to meet the body's demands. The goal of this research is to use machine learning approaches to anticipate which model and classifier will work best for assessing heart failure. The UCI machine learning repository provided access to data from 299 patients with 13 different features. In order to comprehend the data, Explanatory Data Analysis (EDA) was carried out using a variety of methods, including pre-processing, measure of position and distribution, and visualization analysis.

II. Research Plan

The study identified the most effective model and classifier for heart failure prediction using machine learning methods. Three distinct datasets—the original dataset, the imbalanced dataset, and the dataset with significant features chosen using the Random Forest, Mann Whitney U, and Chi-Square feature selection techniques—were used. The models were constructed using a variety of algorithms, including K-nearest neighbor, Random Forest, Support Vector Machine, Naïve Bayes, Logistic Regression, and Neural Networks.

Sample Location:

The dataset that was taken from the UCI machine learning repository was used in the investigation. Using pertinent libraries and the Python programming language, data analysis and model construction were completed.

Sample size: 299 instances with 13 attributes made up the dataset, which included information from heart failure patients. It was decided that the sample size was adequate for machine learning analysis and the development of predictive models for the diagnosis of heart failure.

Thirteen (13) clinical features:
1. age: age of the patient (years)
2. - anaemia: decrease of red blood cells or hemoglobin (boolean)
3. - creatinine phosphokinase (CPK): level of the CPK enzyme in the blood (mcg/L)
4. - diabetes: if the patient has diabetes (boolean)
5. - ejection fraction: percentage of blood leaving the heart at each contraction (percentage)
6. - high blood pressure: if the patient has hypertension (boolean)
7. - platelets: platelets in the blood (kiloplatelets/mL)
8. - sex: woman or man (binary)

9. - serum creatinine: level of serum creatinine in the blood (mg/dL)
10. - serum sodium: level of serum sodium in the blood (mEq/L)
11. - smoking: if the patient smokes or not (boolean)
12. time: follow-up period (days)
13. - [target] death event: if the patient died during the follow-up period (boolean)

III. Result

The mean of each attribute is described in the mean columns of table 1 below. Data is split in half by median values, where 50% of the data lay below and 50% of the data lie above. The majority of the variables in this dataset have similar mean and median values.

The other common measure of spread, variance, is less interpretable than standard deviation, which is the square root of the variance. The mean is the starting point for both variance and standard deviation. Kurtosis quantifies the amount of data in the center and tails of a distribution, whereas skewness quantifies the skewness of a distribution.

The measure spread of each variable is calculated using the minimum and highest values.

Table No 1: Measure Of Location And Spread Numeric Data

Attributes	Mean	Median	Standard deviation	Variance	Minimum	Maximum	Skewness	Kurtosis
Age	60.833893	60.0	11.894809	1.414865	40.00	95.0	0.423062	-0.184871
Creatine - phosphokinase	581.839465	250.0	970.287881	9.414586	23.00	7861.0	4.463110	25.149046
Ejection fraction	38.083612	38.0	11.834841	1.400635	14.00	80.000	0.555383	0.041409
Platelets	263358.029264	262000.0	97804.236869	9.565669	25100.0	850000.0	1.462321	6.209255
Serum-creatinine	1.393880	1.1	1.034510	1.070211	0.5	9.4	4.455996	25.828239
Serum sodium	136.625418	137.0	4.412477	1.946996	113.00	148	-1.048136	4.119712
time	130.2608870	115.0	77.614208	6.023965	4.0000	285.0	0.127803	-1.212048
Anaemia	0.431438	0.000	0.496107	0.246122	0.0000	1.0000	0.278261	-1.935563
Diabetes	0.418060	0.000	0.494067	0.244102	0.0000	1.0000	0.333929	-1.901254
Hugh blood pressure	0.351171	0.000	0.478136	0.228614	0.0000	1.0000	0.626732	-1618076
Sex	0.648829	1.000	0.478136	0.228614	0.0000	1.0000	-0.626732	-1618076
Smoking	0.321070	0.000	0.467670	0.218716	0.0000	1.0000	0.770349	-1.416080
Death Event	0.321070	0.000	0.467670	0.218716	0.0000	1.0000	0.770349	-1.416080

Visualization of Attrributes

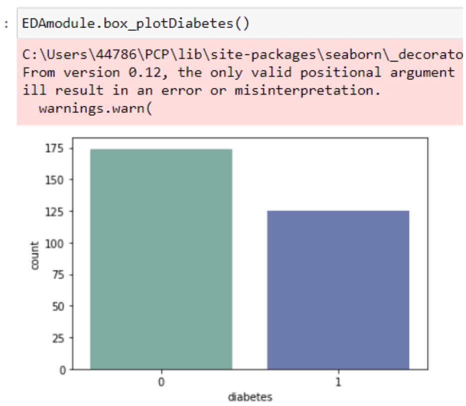


Figure 1. Graph of Diabetes using Bar Plot

It may be deduced from the above graphic that individuals with heart failure without diabetes are roughly 175% more common than those with heart failure with diabetes, who are roughly 125% more common.

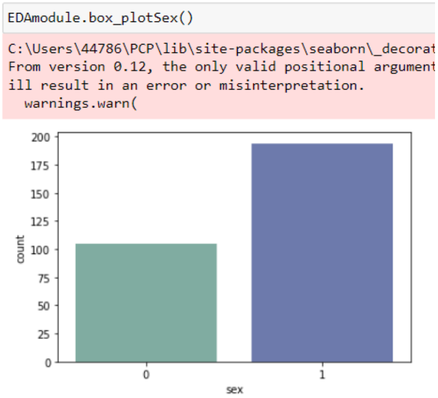


Figure 2. Graph of Sex using Bar Plot

Male = 0, Female = 1, The graph also demonstrates that whereas the rate of heart failure in female patients is 200%, the rate in male patients is only approximately 100%. This graph shows that women are more likely than men to develop heart failure based on the dataset.

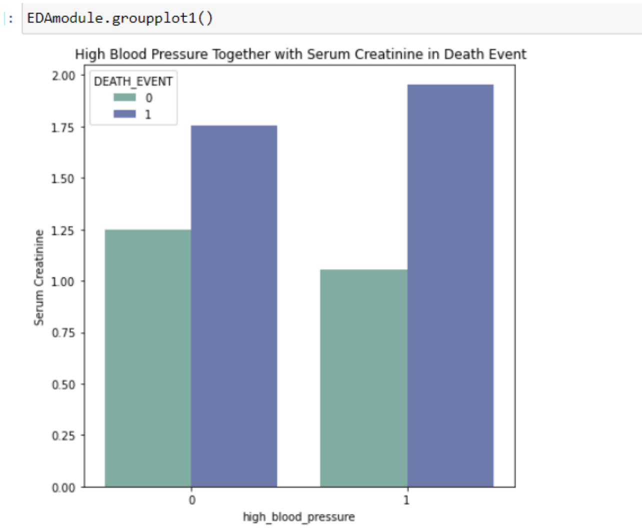


Figure 3. Graph of High Blood Pressure, serum Creatinine with death event

From the graph above, it can be inferred that patients with heart failure who do not have high blood pressure but who have serum creatinine levels of 120 m/gl survive a death event while patients with the same characteristics—no high blood pressure and 175 m/gl of serum creatinine—die in a death event. Patients with high blood pressure and serum creatinine levels of 120 mg/gl escape the death event, but those with high blood pressure and levels of 195 mg/gl perish in the event. According to this graph, people with heart failure who have blood creatinine levels above 121 mg/dl are at risk of passing away, regardless of whether they have high blood pressure or not. See appendix for Further Graph

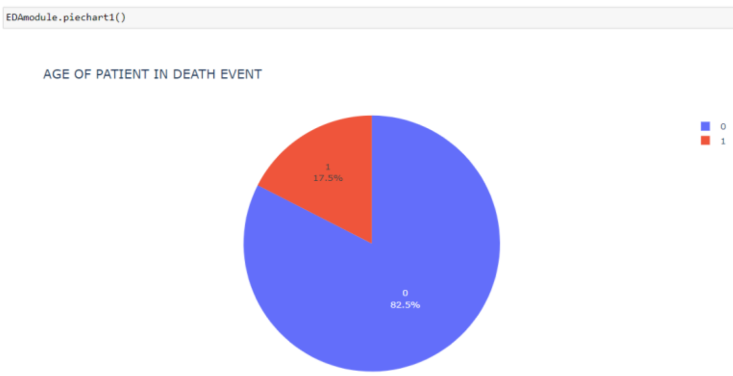


Figure 4. Pie chart of time, age, and death event

This pie chart shows that 82.5% of heart failure patients who spent 32,143 days alive survived a death event, whereas 17.5% of patients who spent 6805 days alive died in a death event. For further graphs, see the appendix.

Conclusion: The explanatory data analysis shows that the dataset has been pre-processed, normalised, and visualised using a bar plot, group bar plot, box plot, correlation graph, and pie chart to analyse the data. There are no missing values in the dataset.

Classification Analysis 1

The optimal model for heart failure analysis in the future can be found or predicted via classification analysis. When building the model, the following algorithm is used. Neural networks, Support Vector Machines, Random Forest, K-nearest Neighbor, Naive Bayes, and Logistic Regression

a) Prior to starting to build the classifier model, the Target variable and Explanatory variables need to be defined as X and Y.
b) By generating training and testing datasets, one can discover how a specific combination of attributes leads to a specific outcome. In order to create all the models, the data was split into 20% test samples and 80% train samples.
c) Create a Classifier Model: In order to comprehend the dataset, the aforementioned techniques will be developed here.
d) "Fitting the model into the training and testing set" refers to using the constructed model for the available training and testing set. For the model to correctly predict outcomes, it must be fitted.

Table No 2: Classification Metrics With Original Data Orignal Dataset

Classifier	Accuracy	Precision	Recall	F1-Score
Naïve Bayes Original data	0.71666	0.57142	0.6000	0.58536
Logistic Regression _Original data	0.83333	0.75000	0.75000	0.75000
SVM_Original data	0.83333	0.77777	0.70000	0.73684
Random Forest_Original data	0.86666	0.80000	0.80000	0.80000
KNN_Original data	0.60000	0.35714	0.25000	0.29411
NN_Original data	0.33333	0.33333	1.00000	0.50000

As can be seen from table 2. the RF model did well on the dataset; its F1-score, accuracy, precision, and recall were all higher than average at 86%, 80%, 80%, and 80%, respectively. The performance of LR, NB, and SVM was excellent. The neural network had biases.

RF, SVM, and LR NB were found to be the most effective algorithms for the built model, and it was recommended that this dataset be used for further study and forecasting.

Classification Analysis II

Investigate Class Imbalance Problem: The imbalance dataset was investigate and Survival Class(0) has the majority class while the Dead Class(1)has minority class.

Figure 4. Graph of Imbalanced Dataset

Sampling Technique using SMOTE: (Opeyemi, 2021) Synthetic minority oversampling Method This methodology also considers the over-sampling of minority classes using artificial data. Instead of oversampling with replacement, provide an example. Equal numbers of synthetic classes are produced by SMOTE for the minority class

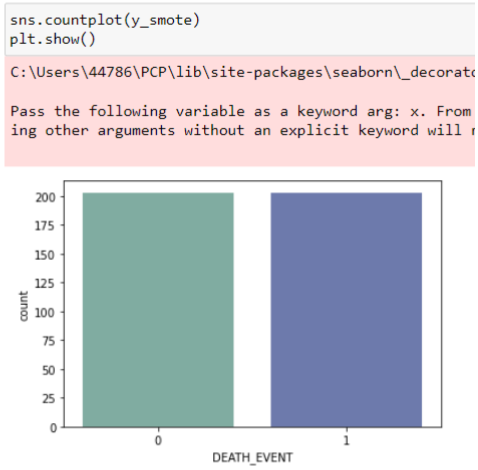


Figure 5: Graph of Balance Data Using SMOTE

After the data was balanced, the SMOTE model was split into training and testing, employing 80% and 20% of the total data. After constructing the classifier model and fitting the SMOTE model, the classification metric was examined below.

Table 3: Classification Metrics With Smote Dataset

CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
NAÏVE BAYE_SMOTE	0.817073	0.842105	0.78048	0.81012
LOGISTIC REGRESSION_SMOTE	0.78048	0.767441	0.80487	0.78571
SVM_SMOTE	0.76829	0.76190	0.78048	0.77108
RANDOM FOREST_SMOTE	0.87804	0.87804	0.87804	0.87804
KNN_SMOTE	0.60975	0.68000	0.41463	0.51515
NN_SMOTE	0.50000	0.50000	1.00000	0.66666

The SMOTE model and the six algorithms that were used on it are shown in Table 3.1. It can be concluded that the SMOTE model works well with Random Forest SMOTE, Naive Baye SMOTE, and Logistic Regression. The SMOTE model's Random Forest approach is the one that is most frequently employed; it performs well in the dataset and has 87% accuracy, 87% recall, 87% precision, and 87% F1 Score. Both Naive Baye and Logistic Regression displayed excellent performance. Neural Network Model, in contrast, has bias due of the model. It might be claimed that Random Forest and Naive Baye can both be used for prediction.

ADASYN (Adaptive Synthetic), an algorithm that creates synthetic data, has the advantages of creating more data and not copying the same minority data.

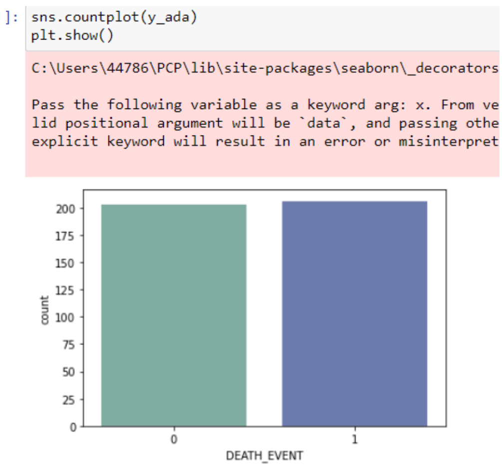


Figure 4: Graph of Balance Data Using ADASYN

Table 4. Classification Metrics With Adasyn Dataset

CLASSIFIER	ACCURACY	PRECISION	RECALL	F1-SCORE
NAÏVE BAYE_ADASYN	0.50000	0.50000	0.48780	0.49382
LOGISTIC REGRESSION_ADASYN	0.80487	0.80487	0.80487	0.83950
SVM_ADASYN	0.78048	0.75555	0.82926	0.79069
RANDOM FOREST_ADASYN	0.86585	0.85714	0.87804	0.86746
KNN_ADASYN	0.56097	0.59259	0.39024	0.47058
NN_ADASYN	0.48780	0.49350	0.92682	0.64406

Table 4 displays the results of the algorithm models that balanced the dataset using ADASYN sampling strategies. With 86% Accuracy, 87% Recall, and 86% F1 Score, it can be concluded that Random Forest fared better and should be considered the best model for this classifier II. This makes it easier to decide that Random Forest Model works better when utilising ADASYN. Additionally, it was shown that the neural network model did not exhibit bias, in contrast to the other classifier models used in Classification 1 and SMOTE.
Additionally, with 80% accuracy, 80% recall, and 83% F1 scores, logistic regression performs better.

Given an unbalanced dataset, SMOTE performs better than ADASYN because each of the six models it uses has at least a 50% accuracy rate.

Based on the classification metrics of the two models, SMOTE is now advised in order to balance the dataset and accommodate algorithms like Random forest, logistic regression, and support vector machine.

Feature Selection of Importance Attributes

The most crucial attributes for this dataset on heart failure have to be chosen for this step. The chi-square test, Mann Whitney U, and random feature selection were advised as three distinct strategies to use while making this choice.

Table 5. Chi-Square Test

s/n	Features	p-values
1.	Creatinine_Phosphokinase	0.000000
2.	Time	0.000000
3.	Platelets	0.000000
4.	Ejection_Fraction	0.000000
5.	Age	0.000000
6.	Serum_Creatinine	0.000034
7.	Serum_Sodium	0.235316
8.	High Blood Pressure	0.340160
9.	Anaemia	0.444804
10.	Diabetes	0.546791
11.	Smoking	0.762509
12.	Sex	0.814814

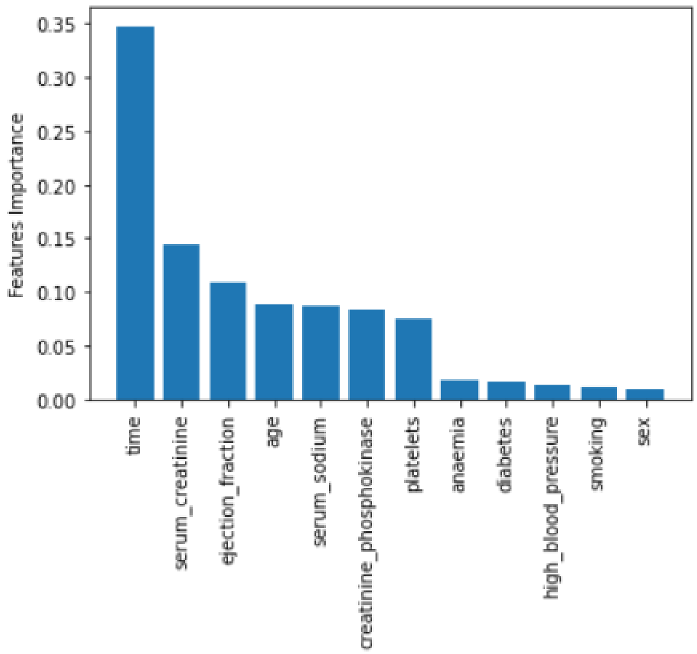
Our P-value, which indicates the significance of those qualities to the dataset, is listed from the chi-square test above in ascending order. Only six features match the requirements for choosing this based on the p value of 0.05 that is provided.

Tables 6 Mann-Whitney U Test

Rank	Features	p-Values (MannWU)
1.	Ejection Fraction	0.000000
2.	Serum Sodium	0.000000
3.	Creatinine Phosphokinase	0.000000
4.	Age	0.000000
5.	Platelets	0.000000
6.	Time	0.000000
7.	Serum creatinine	0.000000
8.	Sex	0.000000
9.	Anaemia	0.000630
10.	Diabetes	0.003672
11.	High Blood Pressure	0.067294
12.	Smoking	0.442066

The Mann-Whitney test We have placed the results of the U test above our P-value in ascending order to indicate the relative importance of each characteristic to the dataset. We may infer that 10 out of the twelve traits, or 95 percent of them, satisfy the requirements for selecting this from the given p value of 0.05.

: (-1.0, 12.0)



Graph 5: Feature Selection Using Random Forest Model

According to the Random Forest feature selection graph, a dataset's relevance is correlated with the value graph's height. It can be inferred that seven attributes satisfy the criteria for the choice.

Finally, seven features—Age, Creatinine Phosphokinase, Ejection Fraction, Platelets, Serum Creatinine, Serum Sodium, and Time have been carefully chosen among the three techniques.

Classification Analysis III

Table 6: Classification Metrics With Important Feature Selection Dataset

VARIABELS	ACCURACY	PRECISION	RECALL	F1 SCORE
Naïve Bayes	0.71666	0.57142	0.6000	0.58536
Logistic Regression	0.83333	0.75000	0.7500	0.75000
SVM	0.80000	0.70000	0.70000	0.70000
Random Forest	0.85000	0.78947	0.75000	0.76923
KNN	0.65000	0.40000	0.10000	0.16000
Neural Network	0.66666	0.0000	0.00000	0.00000

In the classification metrics shown in Table 5.1 above, Random Forest, Logistic Regression, and Support Vector Machine operate best with 85%, 83%, and 80% accuracy, respectively. improved recall and F1 score. With this dataset, bias in neural networks.

Comparison Of Classification Phase

Using various categorization methods, this phase will explain how to improve the various models that were used for the examination of the negative (minority) class. Several metrics, including Accuracy, Recall, and F1-Score, will be used to assess the classification performance.

Accuracy is the overall how often the classifier is correct.

$$\frac{TP + TN}{OVERALL}$$

Recall: $\frac{TP}{TP+FN}$

F1-Score The harmonic means of precision and recall for the minority positive class is used to determine the F1 score, also known as F-measure or balanced F-score, an error metric that assesses model performance. f1 works well with datasets that have imbalances.

$$2 * \frac{Precision * Recall}{Precision + Recall}$$

The accuracy of positively predicted labels is a measure of precision.

$$\frac{TP}{(FP+TP}$$

It is concluded that Classification II was recommended based on the highest performance of the algorithm with the higher value of Accuracy, Precision, and F1 score. This was done after carefully analysing the three classifications used for this dataset: Original dataset classification, Sampling technique model to balance the dataset, and the importance features selection.

Classification II advises analysing the causes and effects of each attribute on the dataset for heart failure.

IV. Discussion

The overall goal of this journal article is to forecast the conditions of heart failure by using machine learning algorithms to data from 299 patient instances with 13 factors. To make sense of the data, the study starts with exploratory data analysis (EDA). Afterwards, models are built using a variety of methods, including Naïve Bayes, Logistic Regression, Support Vector Machine, Random Forest, K-nearest neighbor, and Neural Networks.

The original dataset, the imbalanced dataset, and the dataset with relevant features chosen using feature selection methods such Random Forest, Mann Whitney U, and Chi-Square were all compared as part of the research strategy. The study found that the Classification II model performed better than the others in terms of prediction accuracy and precision through the measurement of several performance indicators, including Accuracy, Recall, and F1-Score.

This study is important because it uses machine learning techniques to determine which model and classifier works best for assessing heart failure scenarios. The study offers insights into enhancing the diagnosis of heart failure by resolving imbalanced data with sampling approaches such as SMOTE and ADASYN, and by finding the most important aspects using feature selection methods.

In conclusion, by proving that machine learning is a useful tool for predicting heart failure, the research presented in this journal article advances the field of medical diagnostics. In order to enhance healthcare outcomes, the study emphasizes the significance of applying advanced data analysis techniques by comparing various classification models and resolving imbalanced data.

V. Conclusion

Based on the performance parameters examined in this study, Random Forest was shown to be the most effective algorithm for predicting heart failure. In order to improve model performance, the significance of feature selection, visualization analysis, and data pre-processing was emphasized. Subsequent studies may concentrate on investigating other machine learning algorithms and methodologies to enhance the heart failure prediction even more.

References

[1]. Davide Chicco & Giuseppe Jurman. (2020, February 3). Machine Learning Can Predict Survival Of Patients With Heart Failure From Serum Creatinine And Ejection Fraction Alone. BMC Medical Informatics And Decision Making, 20(1). Doi:Https://Doi.Org/10.1186/S12911-020-1023-5

[2]. Https://Docs.Scipy.Org/Doc/Scipy/Reference/Generated/Scipy.Stats.Mannwhitneyu.Html#Scipy.Stats.Mannwhitneyu. (N.D.).

[3]. Https://Imbalanced-Learn.Org/Stable/References/Generated/Imblearn.Over_Sampling.ADASYN.Html. (N.D.).

[4]. Https://Scikit-Learn.Org/Stable/Modules/Preprocessing.Html#Preprocessing. (N.D.).

[5]. Opeyemi, A.-A. F. (2021, May 5). Using Ensemble Learning To Fix Imbalanced Data Set. International Journal Of Advances In Engineering And Management, 3(2395-5252), 653-672. Doi:DOI: 10.35629/5252-0305653672