SCHOLARLY REVIEW ONLINE

contact@scholarlyreview.org

SRO: Student Showcase

Winter 2023/2024

About the SRO: Student Showcase

The *SRO: Student Showcase* presents research papers and literature reviews as well as graphic representations, data analysis, and abstracts. This issue of the *SRO* also collects some of the finest finalist papers submitted to our platform.

It is hoped that via this *Showcase* the scholarship on display will reach a wider specialist and non-specialist audience and generate feedback and a scholarly profile for the authors. Every paper in this collection has gone through a rigorous internal review process.

TABLE OF CONTENTS

Application of Machine Learning to Detect Bank Note Fraud

By Allen Ting

AUTHOR BIO

Allen Ting is a high school student who developed a keen interest in mathematics since elementary, which has continued to grow. Recently, he has been exploring the field of machine learning and its underlying theories and models. He's particularly intrigued by the practical applications of machine learning, such as its potential to detect banknote fraud. After looking into the theory behind different types of machine learning models and looking into the alarming impacts of counterfeit currency, Allen has worked to identify the key features that ultimately best differentiate a genuine bank note from a fake one. As a young student, he aspires to contribute meaningfully to the field of machine learning and mathematics in the future, aiming to make a practical and widespread impact.

ABSTRACT

We will discuss how to detect whether a bank note is genuine or counterfeit using machine learning. We first will explain the basic concepts and mathematics behind machine learning. Then, using the binary cross-entropy model and logistic regression, we differentiated between real and counterfeit banknotes from a dataset of banknote dimensions.The results showed that the diagonal length was the most important dimension when classifying the bills, with an error of 0.1546 and accuracy of 0.98, and that inputting the length of the bottom margin and diagonal trained the model to have an error rate under 10%.

INTRODUCTION

Counterfeit bank notes are a problem that potentially affects the nation's economy in a significant manner [3,4]. Although counterfeit currencies are typically produced for personal gain, there also exist instances where counterfeiting money was used as a political weapon against rival nations. When a large number of counterfeit bills are in circulation in the economy, inflation rises due and companies lose monetary value to this artificial increase. Genuine bills lose their true value, and as a result, prices will be marked up, forcing everyday people to spend more money on goods and services. Therefore, it is important to create a machine that can identify the legitimacy of a banknote before it is lost in circulation.

Machine learning, also known as artificial intelligence (AI), is a process that constructs models that perform a certain task without being directly programmed to do so [1,2]. The idea is that machine learning algorithms are trained using datasets, which consist of input data, or features, and corresponding output values, or labels. The algorithms learn from these examples by detecting patterns, extracting relevant features, and generalizing from the data to make predictions on new data.

Machine learning is widely used for data analysis and has numerous applications in many fields. In medicine, machine learning is used to predict diseases and illnesses like heart disease and diabetes before it is too late. Machine learning is used to predict stock prices, create self-driving cars, and suggest products and services based on previous purchases. In the banking industry, machine learning can be used to protect private data, decide loan applications, and detect fraud like counterfeit banknotes.

In this paper, we will train a model that can predict whether a swiss banknote is genuine or counterfeit from a dataset containing 100 genuine and 100 counterfeit notes. The dataset came from Kaggle, a website that has numerous datasets that is free for the public [5]. These datasets can be used to train machine learning models.

DATA PREPROCESSING

The dataset consists of 200 Swiss banknotes, 100 that are genuine and 100 that are counterfeit. Although the banknotes in the dataset are real banknotes, the dataset is synthesized and is not extracted from a real sample. An example of the data in the dataset is shown in table 1, which show the dimensions of the first five banknotes and if they are genuine or counterfeit.

Counterfeit	Length	Left	Right	Bottom	Top	Diagonal
0	214.8	131.0	131.1	9.0	9.7	141.0
0	214.6	129.7	129.7	8.1	9.5	141.7
0	214.8	129.7	129.7	8.7	9.6	142.2
0	214.8	129.7	129.7	7.5	10.4	141.0
0	215.0	129.6	129.6	10.4	7.7	141.8
...						

Table 1: Data for the first five bills in the dataset

The dataset contains 6 features, or characteristics, and 1 label, or output. The 6 features are the following:

x_1 : The length of the bill (Length)
x_2 : The height of the left of the bill (Left)
x_3 : The height of the right of the bill (Right)
x_4 : The height of the bottom margin of the bill (Bottom)
x_5 : The height of the top margin of the bill (Top)
x_6 : The length of the diagonal of the bill (Diagonal)

Figure 1 below shows the measurements for each feature on the Swiss bank note.

Figure 1: Labeled banknote measurements

All measurements are measured in millimeters (mm). The label “counterfeit” states whether the bill with the specified features is genuine (0) or counterfeit (1).

There is no data missing in the 200 banknotes, meaning that it is not necessary to clean the dataset, but we improved the performance of the machine learning algorithm by scaling the data values in the dataset before feeding it into the algorithm. Scaling the data transforms each data value to a number between -1 and 1, thereby reducing the range of the values of the dataset and making it easier for the model to process the information.

TRAINING AND VALIDATION SET

To train and test the model, the data is split into two sets: a training set and a validation set. The training set consists of 75% of the data, while the other 25% of the data is assigned to the validation set. The training set will be used to train and develop the model, and the machine must have both the features (length, left, right, etc.) and the classification (counterfeit or genuine). To measure the accuracy of the model, the validation set will be fed into it without the classification, and the model will then classify whether or not the bill is counterfeit based on the measurements of the features for each bill. It is important to not reveal the classification in the validation set so that we can compute the error for the trained model. If the error is small, then the model will likely make better predictions.

LOGISTIC REGRESSION

One type of machine learning technique for supervised learning problems, the form of machine learning which uses labeled input and output data instead of raw data, is logistic regression. This method only works for binary classification problems. In binary classification problems, the labels, or result, only have two results: 1 or 0. In our dataset, if the result is 1, then the bill is counterfeit. If the result is 0, then the bill is genuine.

Logistic regression is a special case of linear regression because instead of using a line to predict the classification, a logistic function is used instead. This logistic function is also referred to as the activation function. In this paper, the sigmoid function will be the activation function for logistic regression, which is shown in figure 2.

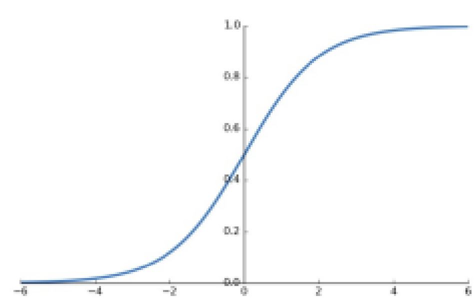


Figure 2: Sigmoid function graph

There are a few properties of the sigmoid function. To begin with, $\sigma(x)$ is an increasing function. Furthermore, $0 < \sigma(x) < 1$ for all x, $\sigma(x)$ becomes arbitrarily close to 0 as x becomes large in absolute value but negative, $\sigma(x)$ becomes arbitrarily close to 1 as x increases, and $\sigma(0) = 0.5$.

In Logistic Regression, the prediction, ˆy, is of the form

$$\hat{y} = \sigma(w_1x_1 + w_2x_2 + \ldots + w_kx_k + b)$$

where σ is the sigmoid function, k is the number of features in the data set, and $w_1, w_2, \ldots, w_k$, b are parameters that minimize the cross entropy error on the training set. In order to understand how to determine the values of $w_1, w_2, \ldots, w_k$, b, we must understand how to calculate the cross entropy error.

BINARY CROSS ENTROPY ERROR

Binary Cross Entropy is defined as a loss function that measures the difference between the predicted probabilities of a binary classification model and the actual binary labels of the data.

Let $y_1, y_2, \ldots, y_n$ be the labels, out outputs for a set of examples. Let $\hat{y}_1, \hat{y}_2, \ldots \hat{y}_n$ be the model's respective predictions for each example. Because this is a binary classification, the labels y_i must be equal to 0 or 1. However, the predictions $\hat{y}_i$ can be values such that $0 < \hat{y}_i < 1$. The cross entropy error is defined as followed:

$$J = \frac{-1}{n}\sum_{i=1}^{n} y_i \log(\hat{y_i}) + (1 - y_i)\log(1 - \hat{y_i})$$

Note that the closer ˆy is to y, the smaller the error. If ˆy = y, then we can observe that the error is 0. The values of $w_1, w_2, \ldots, w_k$, b are chosen to minimize the values of ˆy, thereby minimizing the value of J.

These parameters that are selected are those that make mean binary crosentropy error on the training set as small as possible, and this is done through existing and available libraries like Keras. It is not necessary to dive deeper into the algorithms used to find the parameters that make the error as small as possible.

DATASET ANALYSIS

We apply logistic regression to our swiss bank notes dataset to train the model to predict whether or not a banknote is genuine or not given its dimensions.

```
model = 0
model = Sequential()
#model.add(Dense(1,activation='relu'))
model.add(Dense(1,activation='sigmoid'))
model.compile(loss='binary_crossentropy')
model.fit(X_train_scaled,y_train,epochs=1000,verbose=0)
J_list = model.history.history['loss']
plt.plot(J_list)
```

Figure 3: Binary Cross Entropy Graph Code

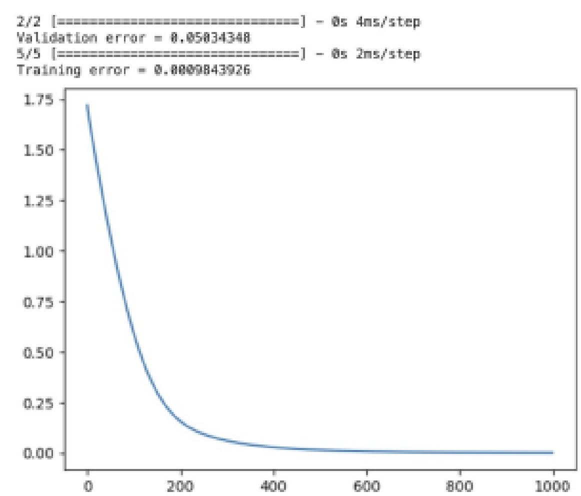


Figure 4: Graph of the error of the model

Figure 3 shows the code used to train the model. In the code, the model reiterates through the scaled training set 1000 times. The number of iterations or cycles the model completes is known as the number of epochs. The model then fits such that it minimizes the error generated from the training set. Figure 4 shows a graph that shows the relationship between the number of epochs and the training error of the model. As the graph flattens out, we stopped cycling through the dataset to prevent overfitting. Overfitting occurs when the model learns the training data too well, including slight variations and noise in the training dataset. When a model overfits the training dataset, it fails to generalize the new data present in the validation set.

```
y_train_hat = model.predict(X_train_scaled)
print('Training error =',bce(y_train.reshape(-1,1),y_train_hat).numpy())

y_val_hat = model.predict(X_val_scaled)
print('Validation error =',bce(y_val.reshape(-1,1),y_val_hat).numpy())
5/5 [==============================] - 0s 2ms/step
Training error = 0.000253629
2/2 [==============================] - 0s 3ms/step
Validation error = 0.060471926
```

Figure 5: Training and validation error

```
y_hat_cat = 1*(model.predict(X_val_scaled) > 0.5)
print(classification_report(y_val,y_hat_cat))
2/2 [==============================] - 0s 5ms/step
              precision    recall  f1-score   support

           0       1.00      0.96      0.98        27
           1       0.96      1.00      0.98        23

    accuracy                           0.98        50
   macro avg       0.98      0.98      0.98        50
weighted avg       0.98      0.98      0.98        50
```

Figure 6: Percent accuracy of the model

The biocrossentropy error when using all the features in the dataset (length, left, right, top, bottom, diagonal) is 0.000253629. On the validation set, the model was able to accurately classify 100% of the genuine bank notes and 96% of the counterfeit bank notes. Therefore, in this given validation set, the model accurately classified 98% of the bank notes in the validation set. The results are shown in figure 5 and figure 6, where figure 5 shows both the training and validation error, and figure 6 shows the accuracy of the model in the chart.

FURTHER INVESTIGATION ON THE FEATURES

To see which features are the most important when training the model, we can isolate each feature and train the model using one feature at a time. Then, by checking the accuracy on the validation set on all 6 features, we can identify which one is the most significant when classifying the bank note.

To do this, we replace the input with one feature in the code and run the code again as is.

```
y = df['conterfeit'].values
#In the previous code, X = df[['Length','Left','Right','Bottom','Top','Diagonal']].values
X = df[['Length']].values
```

Figure 7: Code for one feature

Figure 7 shows the new code used to determine the significance of each feature. By setting the variable X to input one feature at a time instead of all six features, running the same code as the one used on the dataset with all six features produces the accuracy and error for the single feature that was used as the input.

```
2/2 [==============================] - 0s 3ms/step
Validation error = 0.6721552
5/5 [==============================] - 0s 2ms/step
Training error = 0.67482054
```

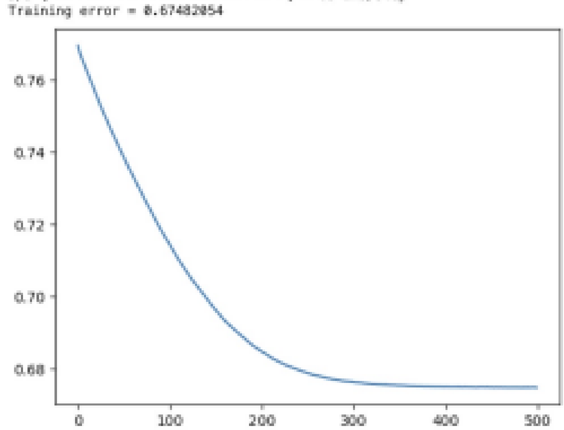


Figure 8: Results after only giving the note's length in mm

Figure 8 shows the error of the model after training it with only the length of the bank notes with an epoch number of 500. After running the code, we get a training error of approximately 0.6748 and a validation error of approximately 0.6725.

This process was repeated for the five other features: Left, Right, Bottom, Top, and Diagonal. Each feature was used to train the model and the training error, validation error, and accuracy for each respective model is shown in table 2.

Feature	Training Error	Validation Error	Accuracy
Length	0.6748028	0.6722918	0.62
Left	0.57226825	0.50991535	0.78
Right	0.49197468	0.46006912	0.82
Bottom	0.27800325	0.3105967	0.92
Top	0.47173423	0.44457167	0.72
Diagonal	0.025608378	0.15463987	0.98

Table 2: Training error, validation error, and accuracy for each feature

These values can be confirmed with a heatmap shown in figure 9 of the correlations between a feature and the label "counterfeit".

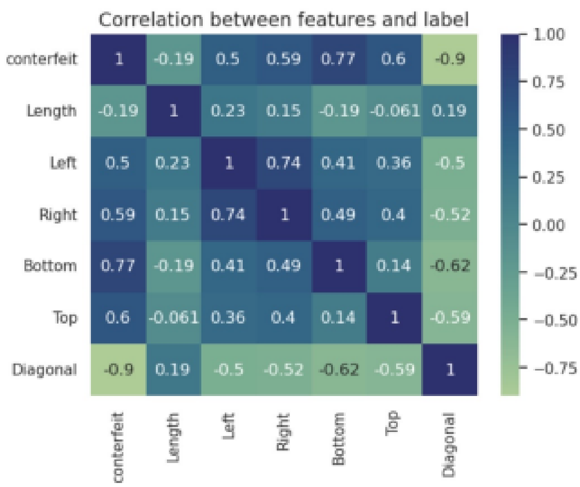


Figure 9: Heat map of the correlation between each feature and the label

The closer the correlation is to -1 or 1, the more related the two variables are. Thus, if the correlation between a feature and "counterfeit" is close to -1 or 1, then the feature is significant when determining whether a bank note is counterfeit or genuine.

Features such as "diagonal" and "bottom" have a high set accuracy (0.98 and 0.92 respectively) and have a low training and validation error compared to the other features. However, their error is high compared to the error we achieved from inputting all the features into the model. We reduced the validation error without inputting all the features by inputting different combinations of pairs of features into the model. To predict which pairs work well together, we use the heatmap and choose the top four features with the greatest absolute correlation value between "counterfeit," pair them up with each other, and feed the two labels into the model to train it.

From Table 3 below, we can observe that the validation error when only diagonal and bottom measurements are inputted are very similar to the validation error when all the features are inputted. Therefore, it is not necessary to use all of the features in the machine learning model. Instead, we only need the dimensions of the bank note's diagonal and bottom to achieve similar results.

Features	Training Error	Validation Error	Accuracy
Diagonal, Bottom	0.008380662	0.062047623	0.98
Diagonal, Top	0.0068245945	0.14657234	0.98
Diagonal, Right	0.009925004	0.13944182	0.98
Bottom, Top	0.04957412	0.088041	0.96
Bottom, Right	0.23715912	0.24233983	0.92
Right, Top	0.36362162	0.32460323	0.82
Right, Left	0.49034384	0.45351523	0.82

Table 3: Training error, validation error, and accuracy for the given pairs of features

APPLICATION OF ML ON BANK NOTE FRAUD

The results show that the logistic regression algorithm is able to accurately predict and classify whether or not a bank note is counterfeit by looking for recurring trends in the dataset. In the end, the model accurately classified 98% of the banknotes correctly with two features, the diagonal dimension and bottom margin dimension with a validation error of 0.062. The high accuracy and low error shows that the model has the potential to classify other sets of bank notes with high accuracy and low error.

Since the model also only needed two features instead of all six, the model may not need many dimensions of the banknotes in future datasets if there is a high correlation between one specific dimension and the classification. Requiring fewer features makes data collection less time consuming and expensive.

REFERENCE

Brown, S. (n.d.). Machine learning, explained. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained (accessed: 06.21.2023).

Francis, R. F. A. (n.d.). A brief history of currency counterfeiting. https://www.rba.gov.au/publications/bulletin/2019/sep/a-brief-history-of-currency-counterfeiting.html (accessed: 06.23.2023).

IBM. (n.d.). What is machine learning? https://www.ibm.com/topics/machine learning (accessed: 06.29.2023).

INTERPOL. (n.d.). Counterfeit currency [fact sheet]. https://www.interpol.int/content/download/10475/file/Counterfeit$%5C%$20currency.pdf (accessed: 06.19.2023).

Riedwyl, B. F. H. (n.d.). Multivariate statistics: A practical approach. london: Chapman hall, tables 1.1 and 1.2, pp. 5-8. https://www.kaggle.com/datasets/chrizzles/swiss-banknote-conterfeit-detection (accessed: 06.10.2023).

Following the Most Suitable Path to Democracy - Philippine Scenario: Democracy to Demo-crazy

By Kaixuan Jin

AUTHOR BIO

Kaixuan Jin is a student of Grade 10 from the Renmin University Affiliated High School International Curriculum Center, Beijing, P.R. China. He is delighted in economics, politics, history and classical literature. He enjoys reading political, historical articles and journals, and analyzes the relationship between political events and the economic aspects. Moreover, he hopes that every country could develop rapidly and healthily based on its unique domestic and international circumstances. Throughout the research process, he has strengthened the courage and dedication to overcome the academic difficulties.

ABSTRACT

The Republic of the Philippines gained independence from the US after World War II in 1946 and boosted brilliant economic status in Asia till the 1960s with the aid and investment from the US. However, the booming economic situation deteriorated which aroused the arguments as compared with other newly founded Southeast Asian countries. Many scholars noticed that the political system played an important role in economic development, while the establishment of democratic system was incapable of safeguarding the healthy development of the economy. As per Samuel Huntington, the performance of democracy outweighed the mere establishment of democratic system (Huntington, 1968). In the Philippines scenario of pursuing the development of democratic system, the performance of democracy was less than satisfactory. The shifts of presidency in 1986 and 2001 driven by two revolutionary movements had actually aroused the political discontinuity and economic instability. Therefore, the paper tries to testify that the performance of democracy affects economic development significantly by analyzing the adverse effect of its political discontinuity on the economy. Through the case study of the Philippines and comparative case study with Singapore in the aspects of national capacity and democratic accountability, the paper substantiates the details of the democratic system and makes an effective response to the prevailing core subject of political institution and economic development, which hopefully might be worthy of reference for other developing countries in the institutional establishment.

Keywords: Philippines, political discontinuity, economic performance, democracy, national capacity, democratic accountability.

INTRODUCTION

The Republic of the Philippines gained independence from the US after World War II in 1946, till when the US had colonized it for around 50 years and had set governmental and educational systems for it. After independence, the Philippines possessed brilliant economic status in Asia till the 1960s with the help and investment of the US. However, the booming economic situation deteriorated which aroused the arguments as compared with other newly founded Southeast Asian countries after World War II.

By the research of Philippines' economic data and the presidential election, it was worth mentioning that the GDP decreased in some years of election. Why the shift of president and the government could lead to the downside momentum for economic development? Could the discontinuity of politics be the reason?

There were two unique revolutionary movements in the Philippines which overthrew the presidents of Ferdinand Marcos in 1986 and Joseph Estrada in 2001 respectively. The national capacity and democratic accountability would be used to study the cause of the discontinuity in politics in the Philippines. Furthermore, Singapore would be chosen as the comparative country in this research.

LITERATURE REVIEW

The IMF conducted a study on "How Does Political Instability Affect Economic Growth?" in 2011 (Ari Aisen, F. J. V., 2016), covering up to 169 countries, and 5-year periods from 1960 to 2004. The results showed that higher degrees of political instability were associated with lower GDP and GDP per capita growth rates, therefore were harmful to economic growth.

The political stability and economic growth are deeply interconnected, as per Zahid Hussain (2014) illustrated. "The uncertainty associated with an unstable political environment may reduce investment and the pace of economic development. And, poor economic performance may lead to government collapse and political unrest." Some scholars also mentioned that the politics discontinuity showed the propensity of a government collapse by severe conflicts or competition between various political parties.

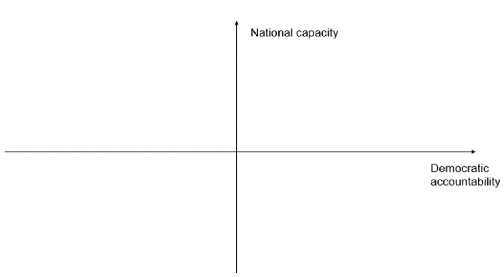
In the book "Politics in the Art of the Possible" (2022), Liu Yu mentioned that politics had a fundamental impact on the fate of a country, especially its stability. Indeed, political continuity was important for economic growth. Liu Yu studied several typical Asian countries as India, Korea and Vietnam in the book, with the purpose to inspire new consideration for other countries of similar geographical location and historical backgrounds.

The author targets the Philippines as the research country which gained independence after World War II and shared some similarities with the countries mentioned by Liu Yu. The independent variable is political instability caused by the shift of the presidency and its cabinet. The dependent variable is economic growth. The interrelation of these two variables will be illustrated by charts. Every time when the independent variable exists, it will lead to the decreasing curve in the economic growth, presenting a negative correlation function, which means that one grows, the other may fall.

The performance of democracy could be propelled by national capacity and the democratic accountability. The author uses the coordinate system with X and Y axis to demonstrate the correlation between national capacity and democratic accountability, with the knowledge gained from the book written by Liu Yu. The X-axis represents the democratic accountability and Y-axis represents national capacity. It is easy to observe the situation of democracy on the graph. More details will be presented in the research finding.

METHODOLOGY

In this essay, the author utilized the methods of text research and qualitative analysis. The book of "Politics in the Art of the Possible" written by Liu Yu, was scrutinized to obtain general knowledge of comparative politics by means of analyzing the correlation between national capacity and democratic accountability. Besides, many academic research papers and journals on political systems, economic development and history had been studied. Furthermore, the author used the method of qualitative analysis to clarify the reasons for economic decline in the Philippines with the data from the World Bank website. Last but not the least, the author drew a coordinate system with X-axis of democratic accountability and Y-axis of national capacity. It is a convenient method to mark the position on the quadrant which indicates the situation of political and democratic development of the countries.



To deepen the understanding of the political discontinuity in the Philippines, the analysis was undertaken for the correlation between national capacity and democratic accountability. Two shifts of presidency were taken into account. First, Ferdinand Marcos failed the reelection in 1986 as a result of the first "People's Power Revolution" and was forced to exile in Hawaii. His opponent Corazon Aquino was elected as the eleventh president of the Philippines. Through the 21-year-tenure, Marcos was overthrown eventually and blamed for running a corrupt, undemocratic regime as a dictator. Both national capacity and democracy were at low level, thus detrimentally affecting the Philippines' economy. Next, the mass elected thirteenth president Joseph Estrada was impeached for corruption of millions of dollars. He was ousted in his third year of tenure in 2001 by the second People's Power Resolution and his vice President Gloria Arroyo ascended to the presidency. This shift of presidency was regarded as the conflicts between elite groups by the outside viewers. The national capacity had little improvement while the mass election pushed forward the democratic process.

FINDINGS

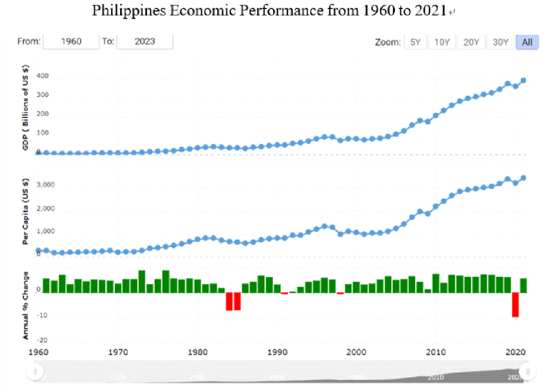
Overall Economic Development in the Philippines

After the independence in 1946, with the aid and investment from the US, the economy in the Philippines had recovered quickly and reached the level before the War, especially in the sector of industry, agriculture and finance (Kästle). In the late 1950s, many multinational companies had transferred their headquarters to Manila for the promising economic prosperity. And Manila was then called the "Mini New York". The GDP per capita in 1961 had reached $278, only next to Japan in East Asia.

However, the boom did not last long when the US shifted its strategic center from Philippines to Japan and Korea, which led to the relocation of the aid and professional personnel. The Philippines' economy had to face the uncertainties because its national economic sectors hadn't developed solidly. The Philippines' government tried to find ways to rebuild its economy in the agricultural sector and labor market and achieved a short period of prosperity starting from the late 1960s.

In the 1970s, the Philippines economy relied heavily on foreign debt, with the ceiling height of 10.5 billion US dollars, more than half of its total GDP. The economy in the 1970s and 1980s came into stagnation with low growth rate and high inflation rate. The Asian Financial Crisis in 1997 hit the economy seriously and caused further deterioration. Since 2010, the economy has stepped on the healthy track of development. The overall economic performance in the Philippines from 1960 to 2021 is revealed in the chart below.

Philippines Economic Performance from 1960 to 2021

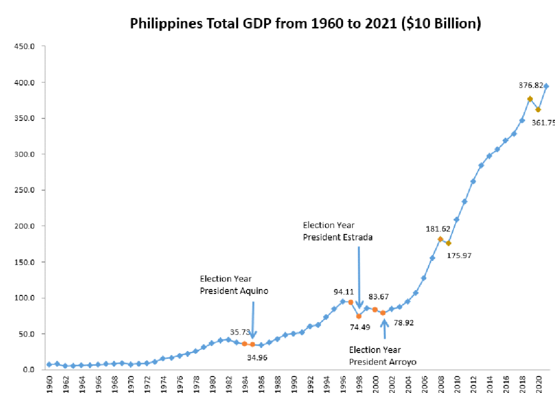


Data Source: World Bank

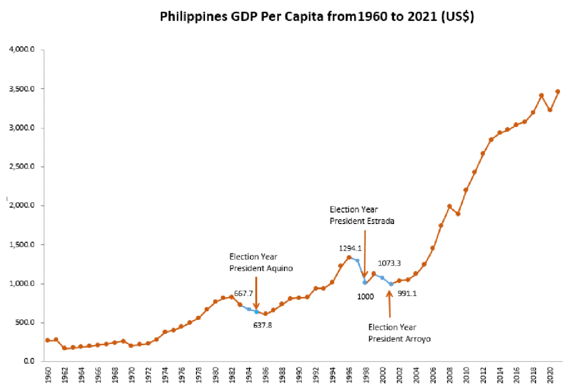
Analysis on the National Capacity and Democratic Accountability in the Philippines

The concept of national capacity was developed by political scientists and economists, mainly referring to "the ability of a state to collect taxes, enforce law and order, and provide public goods" (Lindsey, 2021). While democratic accountability in the political realm often involves "a relationship of representation" (Kitschelt, 2011), by means of "electoral processes and parliamentary oversight, as well as reviews by supreme audit institutions, investigative journalism and public demonstrations" ("Democratic Accountability"). The degree of these two indicators will influence the structure of the country's political system and its mode of democracy, thus exerting positive or negative impact on its economy.

Going through the chart of Philippines overall economic performance in terms of GDP and GDP per capita from 1960 to 2021, five downhill curves could be observed during respective periods of 1985-1986, 1997-1998, 2000-2001, 2009-2010 and 2019-2020. The first three periods of decreasing GDP coincidently happened in the election years. The other two were affected by outside factors as subprime mortgage crisis in 2007 and COVID-19 pandemic in 2019. Therefore, the research will focus on the two presidential shifts which were driven by the revolutionary movements in the year of 1986 and 2001, and leave out the down curve in 1997-1998 which should be evaluated by the intertwined influence of the Asian Financial Crisis and the shift of presidency.



Data Source: World Bank



Data Source: World Bank

Presidential Election in 1986

The Presidency of Marcos

Ferdinand Marcos, Nationalist Party leader, was elected in 1965 as the tenth president and remained in power till 1986 ("Ferdinand Marcos", 2019). During his first presidential term, he focused on construction of infrastructure projects and bolstering the country's rice production which were considered necessary and helpful for the recovery from the War. Marcos was reelected in 1969, the first Filipino president to win a second term, but violence and fraud were associated with his campaign. In his second term, economic growth slowed, prosperity faded, the crime rate increased and political unrests were widely spread. Marcos declared martial law on 21 September 1972, and abolished the Constitution for the purpose of refusing to relinquish power. The martial law was lifted on 17 January 1981. During this time, he implemented military control throughout the country, took over the media and put political opponents in jails. Marcos also manipulated a national referendum supporting his presidency indefinitely in 1973. In June 1981, he won presidential reelection for another six years, with his political opponents boycotting the vote. The end of the Marcos era occurred when his chief political rival, Liberal Party leader Benigno Aquino, who had been jailed by Marcos for eight years, was assassinated on his arrival in Manila on 21 August 1983. Marcos and his wife were blamed for the assassination. In this result, the urban wealthy and middle class, core supporters of Marcos, began to impeach him for enriching his personal fortune via crony capitalism, monopolies and overseas investments that violated the law. To quiet the opposition and reassert his power, Marcos called for special presidential election in 1986, a year before the end of his current six-year term. Corazon Aquino, the widow of Benigno, became the presidential candidate of the opposition. Marcos managed to defeat Aquino and retained the presidency, but his victory was regarded as fraudulent. Thousands upon thousands of citizens took to the streets to support a non-violent military rebellion known as "People's Power Revolution" against Marcos. He was ousted on the date of inauguration on 25 February 1986 and Aquino was brought to power. Marcos and his family were forced to flee into exile in Hawaii with a huge amount of money embezzled from the Philippines.

Marcos' Crony Capitalism

During his presidency, Marcos imposed "crony capitalism", by which private businesses were seized by the government and handed over to friends and relatives of government members, later leading to economic instability. Take his wife, Imelda, as an example. She had been appointed as the mayor of Manila since 1975. In 1978, she was appointed unofficially as the successor of the president. In 1979, she was appointed as the chairman of the cabinet meeting composed of ministers, and became the second most powerful person in the Philippines. Their relatives and close friends were also entitled with senior office posts and took the privilege for seeking benefits of their own. Marcos and his wife were criticized for crony capitalism both politically and economically for which led to the monopolies and the income gap between the rich and the poor.

The Presidency of Aquino

On 25 February 1986, as a result of the "People's Power Revolution", Corazon Aquino became the first female president of the Philippines (Szczepanski, 2020). She restored democracy to the country, promulgated a new constitution which forbade a second term of presidency, and served until 1992.
Presidency for Aquino was not entirely smooth. She pledged agrarian reform and land redistribution, but she could hardly fulfill her promise as opposed by landlord classes. Marcos' supporters staged many coup attempts against Corazon Aquino during her tenure, but she survived with the help of her allies in her unique political style. She refused to run for a second term in 1992 and was praised for setting a good example to abide by the new 1987 Constitution which forbad the second term. She has been described as both the "mother of Philippine democracy" and as the "housewife who led a revolution." (Dungo 2022).

Analysis in Terms of National Capacity and Democratic Accountability

Ferdinand Marcos was the president of the Philippines from 1966 to 1986 before fleeing to the US. He was known for running a corrupt and undemocratic regime, being blamed for his dictatorship, crony capitalism and fraudulence of presidential elections, all of which greatly damaged the democratic accountability. At the same time, he was also criticized for corruption, ineffective governance, unstable social situation and income gap. The national capacity during his tenure was weak. Therefore, the Philippines was located in the third quadrant on the coordinate system.

Since the shift of presidency from Marcos to Aquino by "People Power Revolution", the political instability affected the overall economic performance with a 2.8% drop of GDP in 1985-1986 period.

Shift of Presidency in 2001

Joseph Estrada served as the 13th president of the Philippines from 1998 to 2001 ("Joseph-Estrada", 2023). In 1998 Estrada ran for president, though he was significantly opposed. However, with the support of Imelda Marcos (the widow of former president Ferdinand Marcos) and votes from the country's poor, Estrada won nearly 40 percent of the vote, handily defeating his nearest rival. The margin of victory was the largest in a free election in the history of the Philippines, and Estrada was officially declared president by Congress on 29 May 1998.

The tenure for Estrada as president was short-lived, as a corruption scandal erupted in October 2000 claiming that Estrada had accepted bribes worth millions of dollars. In November the Philippine Senate began an impeachment trial. On 20 January 2001, Estrada was ousted amid mass protests known as the Second People's Power Revolution, and his vice president, Gloria Arroyo, ascended to the presidency. Due to the shift of presidency from Estrada to Arroyo in the 2000-2001 period, the overall economic performance in GDP dropped 5.7%.

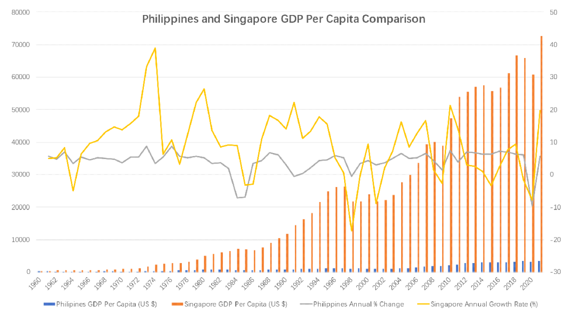
It was ironic that the mass elected president with overwhelming victory was ousted by the People's Power Revolution. Looking through the underlying reasons, the elite groups' conflicts were a common phenomenon in the

Philippines, which were deeply rooted and powerful enough to manipulate the election process. The national capacity had little improvement while the mass election pushed forward the democratic process. Therefore, the Philippines was located in the fourth quadrant on the coordinate system.

The Comparison with Singapore in National Capacity and Democratic Accountability

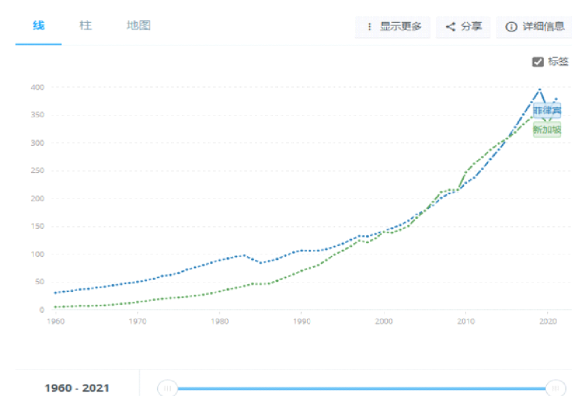
Singapore was founded on 9 August 1965. It has been renowned for the "Singapore model", created by Lee Kuan Yew, the first and most influential Prime Minister from 1959 until 1990. The Singapore model is highly centralized and meritocratic, and the government runs efficiently, rationally and without corruption ("Singapore's Political System", 2019). For more than 50 years, the People's Action Party (PAP) has dominated its politics.

Under the PAP, the government has taken a central role in promoting business, encouraging nationalized companies across a variety of industries. Moreover, the government has kept taxes low and simplified the rules and regulations facilitating private businesses; and it has actively sought out foreign investment, by providing incentives for entrepreneurs seeking to do business in Singapore. Therefore, its economy has developed at a high speed. The charts below show the difference in GDP per capita and GDP from 1960-2021 between Singapore and the Philippines.



Data Source: World Bank

Singapore and Philippines GDP from 1960 to 2021



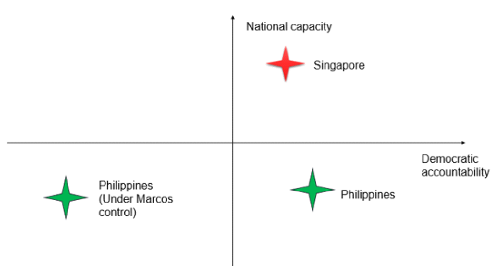
Data source: World Bank

From the above charts, it is notable that the GDP of Singapore has caught up and surpassed that of the Philippines in 2006. The scales of total GDP of these two countries have been equivalent after 2000. However, the GDP per capita has a huge gap. The GDP per capita of Singapore has been almost 20 times higher than that of the Philippines in recent years, which is the indicator reflecting the affluence and economic level of a country. There is no doubt that the momentum of economic development in Singapore is much better than that of the Philippines.

Singapore's political system, which originated from the British, is a Westminster-style democracy. The Constitution provides the framework for Singapore's political system, consisting of three branches of government, the Legislative, the Executive, and the Judiciary which supervised and balanced one and another. Although Singapore is a multi-party nation, the People's Action Party (PAP) has been the dominant force since independence, regularly winning well over 60% of the vote ("Singapore's Political System", 2019).

Therefore, Singapore belongs to the first quadrant of the coordinate system with high national capacity in the aspects of separation of legislative, executive and judiciary powers, efficient governance, excellent incorruptibility, and reasonable democratic accountability on all-citizen-election, effective economic rules and regulations. The coordinate system on the correlation of national capacity and democratic accountability is shown as follows.

The Correlation of National Capacity and Democratic Accountability



DISCUSSION

As illustrated above, the discontinuity in politics really makes a difference in the Philippines. The two shifts of presidency could be regarded as typical examples not only because of the political conflicts steered by the People's Power Revolution, but also because of the consequential drop of economy in terms of total GDP of 2.8% and 5.7% respectively. Further study should be carried to verify the conclusion that the politics discontinuity played negative effects on Philippines' economic development.

What's more, the qualitative analysis was used to locate the position on the coordinate system of national capacity and democratic accountability. It was actually much more complicated for the method of positioning in the system, because many parameters or factors should be taken into consideration. For future study, quantitative methods or data modeling should be adopted to evaluate the relationship of the parameters so as to figure out the exact position in the system.

It is meaningful to visualize the position on the coordinate system of democracy by axis of national capacity and democratic accountability, with the purpose to evaluate the performance of democracy. The position on the coordinate system of democracy synchronizes with the changes of national capacity and democratic accountability.

It is difficult to comment on which mode of democracy is better or worse. However, the conclusion could be drawn that the degree of democracy must be in line with its national conditions. It is extremely difficult for a country to promote democracy while enduring political instability and social under-development. Democracy is like a double-edged sword, if implemented well, it is in favor of promoting the development of the country and the wellbeing of its people. Otherwise, overuse or overstate democracy will bring about demo-crazy, consequently damage the overall development. It is necessary for the emerging countries with similar backgrounds as that of Singapore or the Philippines to take into comprehensive consideration on how to achieve the equilibrium of national capacity and democracy for a better future.

REFERENCE

University, H., & Huntington, S. P. (1968). Political Order in Changing Societies. Yale University Press.

Liu, Y. (2022). Politics in the art of the possible. Guangxi Normal University Press

Hussain, Z. (2014). Can political stability hurt economic growth? World Bank Blogs. Retrieved March 2, 2023, from https://blogs.worldbank.org/endpovertyinsouthasia/can-political-stability-hurt-economic-growth

Ari Aisen, F. J. V. (2016). How does political instability affect economic growth? IMF. Retrieved March 2, 2023, from https://www.imf.org/en/Publications/WP/Issues/2016/12/31/How-Does-Political-Instability-Affect-Economic-Growth-24570

Lindsey, B. (2021). State capacity: What is it, how we lost it, and how to get it back. Retrieved March 2, 2023, from https://www.niskanencenter.org/wp-content/uploads/2021/11/brinkpaper.pdf

Kitschelt, H. (2011). Democratic accountability - duke university. Retrieved March 2, 2023, from https://sites.duke.edu/democracylinkage/files/2014/12/Kitschelt.pdf

Democratic accountability. International IDEA. Retrieved March 2, 2023, from https://www.idea.int/our-work/what-we-do/democratic-accountability

Kästle, K. (n.d.). History of the Philippines. Retrieved March 2, 2023, from https://www.nationsonline.org/oneworld/History/Philippines-history.htm

A& E Networks Television. (2019). Biography.com. Retrieved March 2, 2023, from https://www.biography.com/political-figures/ferdinand-marcos

Dungo, F. (2022). Complete list of presidents of the Philippines - achievements and contributions. Philpad. Retrieved March 2, 2023, from https://philpad.com/presidents-philippines-achievements-contributions/

Szczepanski, K. (2019). Biography of Corazon Aquino, first female president of the Philippines. ThoughtCo. Retrieved March 2, 2023, from https://www.thoughtco.com/corazon-aquino-biography-195652

Encyclopædia Britannica, inc. (2023). Joseph Estrada. Encyclopædia Britannica. Retrieved March 2, 2023, from https://www.britannica.com/biography/Joseph-Estrada

Singapore's political system. StartupDecisions. (2019). Retrieved March 2, 2023, from https://www.startupdecisions.com.sg/blog/singapores-political-system/

Analyzing Morphological Types and Color Indices of Galaxies in Clusters

By Ethan Tang

AUTHOR BIO

Ethan Tang is seventeen years old and is attending his senior year at Westwood High School. He has interests in the field of computer science, astrophysics, and data science, and is passionate about music. At school, he plays the saxophone in the Westwood Band, and he spends his afternoons at marching band rehearsals. At home, he also enjoys playing the piano.

ABSTRACT

Galaxy morphology is important for understanding the formation and evolution of galaxies within a cluster, as morphological type has been attributed to distance and location. The colors of galaxies are also very useful for analyzing their properties. This paper analyzes galaxy morphology, looking at the ratio of Early Type Galaxies to Late Type Galaxies within galaxy clusters at changing redshifts. Furthermore, the color indices of these galaxies are explored as the paper looks into how the colors of galaxies differ at various redshifts. It was found that Early Type Galaxies were dominant in clusters at redshifts less than one, while Late Type Galaxies were dominant in those at redshifts greater than one. Additionally, Early Type and Late Type Galaxies had similar changes in colors from redshifts 0-1, with Late Type Galaxies always appearing more blue than Early Types.

INTRODUCTION

The concepts that this research focuses on help the astronomical community better understand the different galaxy morphological types and their characteristics. Galaxy morphology is useful for understanding the formation as well as evolution of galaxies, and future research on galaxy formation will benefit from our research as well. By analyzing color filter values, we gain more knowledge about stars, as color is an indicator of star formation. The color of galaxies also tells us about the temperature of galaxies and helps us further explore galaxy morphology. By analyzing clusters, we can see how these galaxies interact in a more crowded environment. Redshift helps astronomers compare the distances of objects that are far away. This paper looks at galaxy clusters at varying redshifts, aiming to find patterns about how the properties of galaxies within these clusters change as a function of redshift.

METHOD

This research collected data from three main sources: Literature searches, the SIMBAD database, and SDSS data. SIMBAD proved to be the most impactful out of these three, providing data on galaxy clusters, galaxies within clusters, morphological types, and redshift. To find the other data points, spectroscopic data from the SDSS database was utilized. Using the u and r color magnitudes of each galaxy, the color magnitudes of u-r were compared to the value of 2.2 to determine the morphological type of each galaxy. If the resulting color index value was greater than 2.2, then the galaxy was classified as an Early Type. If it was less than 2.2, it was classified as a Late Type. SDSS provided the necessary data for the redshift, B magnitude, and V magnitude, which were used to analyze the color indices. A variety of software tools aided in analyzing this data through graphs. These tools included Jupyter Notebooks, Numpy, Matplotlib, and Pandas. The research employed Pandas through Jupyter Notebooks to make .csv data usable in graphs. With the .csv file in Jupyter Notebook, Matplotlib was used to create scatter plots comparing the percentages of Early and Late galaxies at our redshift intervals.

RESULTS

It was found that at the lowest redshift bin, Early Type and Late Type Galaxies are almost equal in frequency. As the redshift increases, Early Type Galaxies begin to dominate until galaxy clusters at redshift 1.5. At this point, the frequencies are equal again and from then on, Late Type Galaxies dominate, making up over 90% of the galaxy cluster at the highest redshift observed. Additionally, the color indices of Early Type and Late Type Galaxies followed a very similar pattern as the redshift increased. Initially, the color index value increases, signifying that the galaxies are appearing more red At redshifts 0.75 to 1, this value drops significantly. Overall, the color index values of Late Type Galaxies are less than the color index values of Early Type Galaxies at every redshift bin.

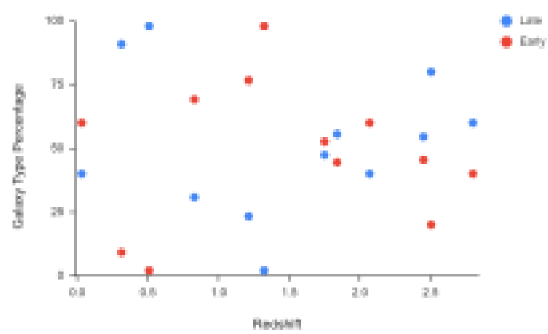


Figure 1. *Percentages of ETGs and LTGs in clusters at different redshifts*

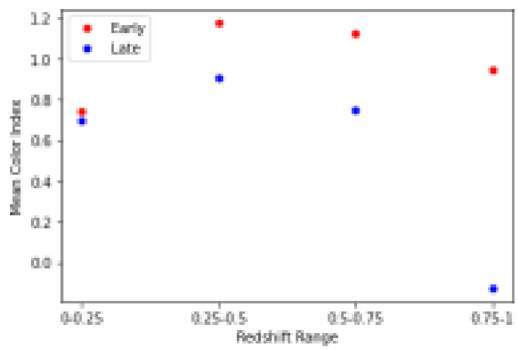


Figure 2. *The average color indices (B-V) of different types of galaxies at different redshifts.*

DISCUSSION

The results obtained from the study were supported by other research papers. Specifically, Durret et al. (2015) and Tamburri et al. (2014) found similar results. One of the graphs included in Durret et al. (2015) was almost identical to the one produced from this research. They only looked at redshifts from 0-1 and concluded that at the lowest redshift, the different morphological types are almost equal in number. Additionally, they found that Early Type Galaxies were dominant at low redshifts, in line with our results. Tamburri et al. (2014) analyzed greater redshifts as well, and they gathered that Late Galaxies dominated overall, which can be seen in our graph at greater redshifts.

CONCLUSION

In this research, through an analysis of galaxy morphology and color indices within galaxy clusters at varying redshifts, it was found that the ratio of Early Type Galaxies to Late Type Galaxies in these clusters changes with redshift. At lower redshifts, the two types of galaxies are roughly equal in frequency, but as redshift increases, Early Type Galaxies begin to dominate, continuing until a redshift of around 1.5, at which point the frequencies equalize again and beyond that, Late Type Galaxies become the dominant type. Additionally, the color indices of both Early and Late Type Galaxies followed similar patterns as redshift increased, increasing and appearing redder at first, but dropping at redshifts between 0.75 and 1. The significance of this finding lies in the fact that it helps improve our understanding of the dynamics of galaxy clusters over time. It not only contributes to our understanding of the evolution of galaxies within clusters, but also gives us insight into the factors shaping the galaxies we observe today.

REFERENCE

Durret, F., Adami, C., Bertin, E., Hao, J., Márquez, I., Martinet, N., Maurogordato, S., Sauvaget, T., Scepi, N., Takey, A., & Ulmer, M. P. (2015, June 10). Galaxy clusters in the SDSS stripe 82 based on photometric redshifts. Astronomy & Astrophysics. https://www.aanda.org/articles/aa/full_html/2015/06/aa25293-14/aa25293-14.html

Tamburri, S., Saracco, P., Longhetti, M., Gargiulo, A., Lonoce, I., & Ciocca, F. (2014, October 30). The population of early-type galaxies: How it evolves with time and how it differs from passive and late-type galaxies. Astronomy & Astrophysics. https://www.aanda.org/articles/aa/full_html/2014/10/aa24040-14/aa24040-14.html

Artificial Intelligence for Oral Squamous Cell Carcinoma Detection

By Raghav Herugu

AUTHOR BIO

As a 10th-grade high school student with a passion for technology, this individual has already gained significant experience in web and app development, UI and UX, AI/ML, and backend engineering. With a strong interest in research, they have even authored two papers in the field of cancer, demonstrating an impressive level of dedication and skill. Driven by a desire to learn and grow in the field of technology, this student is constantly seeking out new opportunities to challenge themselves and expand their knowledge. Whether working on personal projects or collaborating with others, they approach every task with a high level of enthusiasm and a willingness to explore new ideas and techniques. With a proven track record of success in the technology industry and a deep commitment to learning and growth, this student is poised to make a significant impact in the field of technology in the years to come.

ABSTRACT

This research paper investigates the potential of Artificial Intelligence (AI) techniques for the detection of Oral Squamous Cell Carcinoma (OSCC) using a Histopathological Image Dataset. Three different AI models were trained with varying levels of regularization and dropout layers using TensorFlow, an open-source software library for machine learning. The study found that AI-based techniques can achieve high levels of accuracy in the detection of OSCC, with one of the models achieving an accuracy of almost 90%. The use of regularization and dropout layers was found to improve the accuracy of the models. The results of this research demonstrate the potential of AI in improving the accuracy and speed of cancer detection, which could lead to earlier diagnosis and better treatment outcomes for patients. This research has implications for the development of diagnostic tools in the field of medical image analysis.

INTRODUCTION

The field of medical image analysis has undergone significant advancements in recent years, with the emergence of Artificial Intelligence (AI) techniques providing new avenues for detecting and diagnosing various diseases. Oral Squamous Cell Carcinoma (OSCC), a type of oral cancer that arises from the lining of the mouth and throat, is one such disease where early detection and diagnosis is critical to improving patient outcomes. In this context, this research paper aims to investigate the potential of AI techniques for the detection of OSCC using a Histopathological Image Dataset.

Convolutional Neural Networks (CNNs) have become a popular deep learning architecture for image recognition tasks due to their ability to learn spatial features and classify objects with high accuracy (O'Shea, Nash). CNNs utilize a hierarchical structure of multiple layers, including convolutional layers, pooling layers, and fully connected layers, to extract increasingly complex features from the input image.

Over the years, various studies have been conducted to explore the use of AI in medical image analysis. For instance, Chaunzwa et al. (2021) investigated the use of deep learning algorithms for the classification of lung cancer from CT images, while Ozsahin et al. (2023) explored the potential of AI-based techniques for the detection of breast cancer. While these studies have shown promising results, there is a need for further exploration of AI techniques for the detection and diagnosis of other types of cancer, such as OSCC.
The rationale for this research is to improve the accuracy and speed of OSCC detection and diagnosis, which could lead to earlier intervention and better treatment outcomes for patients. The use of AI in medical image analysis has the potential to address the limitations of traditional methods, such as human interpretation of images, which can be subjective and prone to errors. By using a Histopathological Image Dataset and TensorFlow, an open-source software library for machine learning, this research aims to train and evaluate three different AI models with varying levels of regularization and dropout layers, to determine their efficacy in detecting and classifying OSCC cells.

The hypothesis of this research is that AI-based techniques can achieve high levels of accuracy in the detection of OSCC, with the use of regularization and dropout layers improving the accuracy of the models. This hypothesis is based on prior knowledge and observations in the field of medical image analysis, where studies have shown that AI-based techniques can improve the accuracy and speed of disease detection. The prediction is that the use of AI techniques in detecting OSCC will result in improved accuracy and speed, leading to earlier diagnosis and better treatment outcomes for patients.

The overall objective of this research project is to test the efficacy of AI-based techniques for the detection of OSCC using a Histopathological Image Dataset. The independent variable in this study is the AI technology used, with three different models trained and evaluated. The dependent variable is the accuracy and loss of the models, which will be used to evaluate their efficacy in detecting and classifying OSCC cells.

MATERIALS AND METHODS

Histopathological Image Dataset Preparation

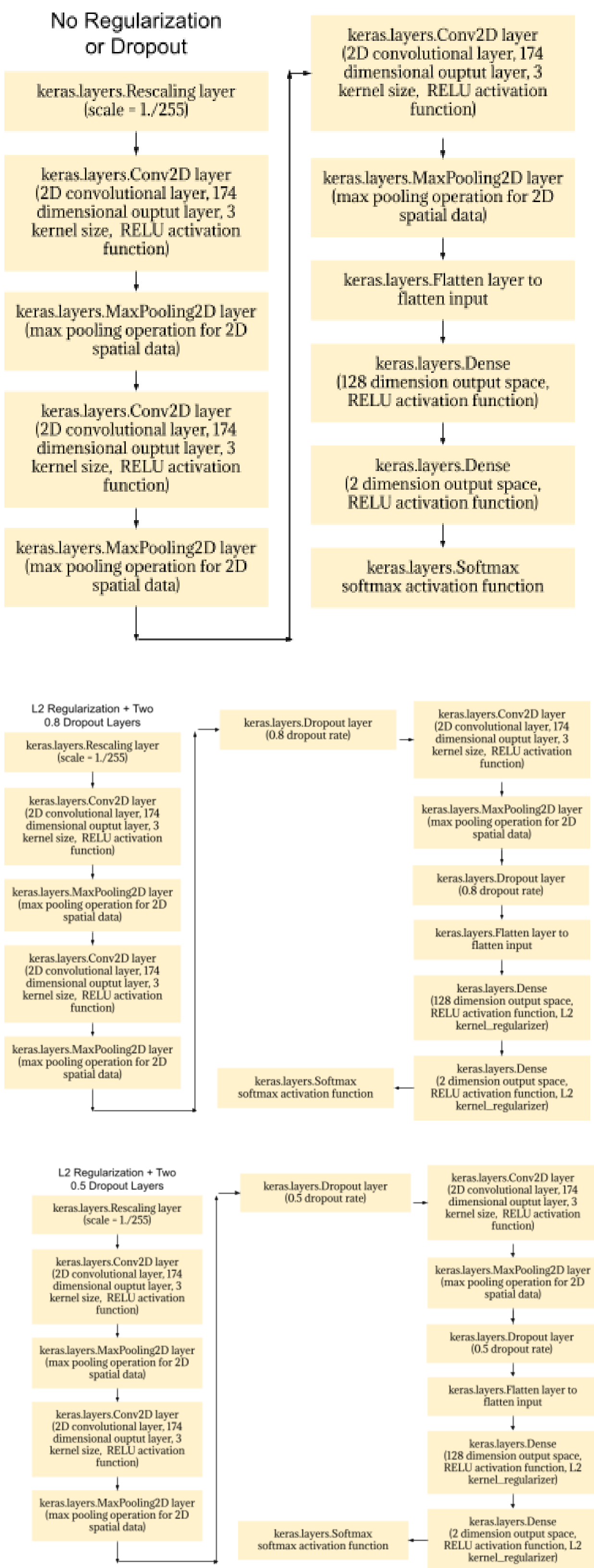
A dataset containing images of Oral Squamous Cell Carcinoma (OSCC) was obtained from a public repository (Rahman). The dataset consisted of images of different magnifications and contained various types of OSCC cells. Data augmentation was performed on this dataset to increase the size of the dataset and improve the accuracy of the models. The augmentation process included rescaling, random flipping, and random cropping. Augmentation was performed on the 400x magnification dataset. This dataset contains 201 Normal Oral Cavity Histopathological (NOCH) images and 495 OSCC images. The data was augmented by adding 36,209 augmented images to the NOCH dataset and 35,002 to the OSCC dataset. This meant that the NOCH dataset had a total of 36,410 images and the OSCC dataset had a total of 35,497 images.

AI Model Training

The choice to develop three different AI models with varying levels of regularization and dropout layers is a significant one for several reasons. Firstly, it allows us to explore the impact of regularization techniques and dropout layers on the model's performance. Regularization methods like L2 regularization help prevent overfitting by adding a penalty term to the loss function based on the magnitudes of the model's weights. Dropout layers, on the other hand, introduce randomness during training by temporarily deactivating a fraction of neurons, reducing co-dependency among them and improving generalization.

The first model, which lacked any regularization or dropout layers, serves as our baseline model. It gives us insight into the model's performance without any regularization or dropout-induced constraints, which can be essential for understanding how these techniques influence the model's behavior. The second model, with L2 regularization and 2 Dropout layers of 0.5, demonstrates a moderate level of regularization and dropout, allowing us to observe the impact of these techniques on mitigating overfitting and model generalization. In contrast, the third model, with L2 regularization and 2 Dropout layers of 0.8, represents a more aggressive application of these techniques. This model's significance lies in showing the potential trade-off between regularization strength and model accuracy. While stronger regularization can enhance generalization, it might also risk underfitting.

Additionally, the hardware specifications of the computer used for training, featuring a 12th Gen Intel(R) Core(TM) i7 - 12650H processor, 32 GB of RAM, and an NVIDIA GeForce RTX 3070 laptop graphics card, play a crucial role in the training process. These high-performance components facilitate faster model convergence and overall efficiency during training, reducing the time and resources required for experimentation. The significance of this hardware setup lies in its ability to support the efficient training of multiple models with various configurations, enabling a comprehensive analysis of the impact of regularization and dropout layers on the AI models' performance.



Evaluation

The trained models were evaluated based on their ability to accurately detect and classify cancerous cells. The evaluation was performed using a separate test dataset that was not used during the training process. The test dataset contained 24 NOCH images and 50 OSCC images. The accuracy and loss were calculated for each model. Each model ran with 100 epochs, which each took around 17 and a half hours. Loss and accuracy curves were graphed. The validation dataset used 75% of the images available and the batch size for these models were 32.

Ethical Considerations

This study used publicly available data, and thus ethical approval was not required. However, all procedures were performed in accordance with the relevant guidelines and regulations.

RESULTS

Figure 4: *Calculated Data Table of Test Accuracy and Loss of Each Group*

	Test Accuracy (%)	Test Loss
No Regularization or Dropout	81.08%	3.4668
L2 Regularization + Two 0.5 Dropout Layers	89.19%	0.7172
L2 Regularization + Two 0.8 Dropout Layers	85.14%	0.5928

Figure 5: Training and Validation Metrics for No Regularization or Dropout

Figure 6: Training and Validation Metrics for L2 Regularization + Two 0.5 Dropout Layers

Figure 7: Training and Validation Metrics for L2 Regularization + Two 0.8 Dropout Layers

DISCUSSION

Summary of Results and Conclusions

The results of this study demonstrate that using regularization and dropout techniques can significantly improve the accuracy of our AI model for detecting oral squamous cell carcinoma. In particular, we found that using L2 regularization and two 0.5 dropout layers resulted in the highest test accuracy of 89.19%, compared to 81.08% without any regularization or dropout, and 85.14% with L2 regularization and two 0.8 dropout layers.

These findings suggest that overfitting was a significant issue in our initial model, and that regularization and dropout techniques were effective in addressing this problem.
Regularization can help prevent overfitting by adding a penalty term to the loss function, which encourages the model to use smaller weights and avoid relying too heavily on any one input feature. Dropout layers work by randomly dropping out some of the neurons during training, which can help prevent co-adaptation of neurons and reduce overfitting.

It is important to note that although L2 regularization and two 0.5 dropout layers resulted in the highest accuracy, the model with L2 regularization and two 0.8 dropout layers still performed better than the model without any regularization or dropout. This suggests that even modest levels of regularization and dropout can be effective in improving the performance of the model.

The results suggest that using regularization and dropout techniques can significantly improve the accuracy of an AI model for detecting oral squamous cell carcinoma. In particular, using L2 regularization and two 0.5 dropout layers resulted in the highest accuracy. These findings have important implications for the development of AI models for medical image analysis, as overfitting is a common challenge in this field. By incorporating these techniques, it may be possible to improve the accuracy and reliability of AI models for detecting cancer and other medical conditions.

Future Studies and Research Impact

The present study provides valuable insights into the use of regularization and dropout techniques for enhancing the accuracy of AI models for oral squamous cell carcinoma detection. However, several avenues for future research can be pursued to build on our findings and to advance the field.

One possible direction for future research is the optimization of the hyperparameters of the regularization and dropout techniques. While we used a fixed set of hyperparameters for L2 regularization and dropout layers, other combinations might lead to higher accuracy. A systematic exploration of the hyperparameter space could help identify the most effective combinations for different types of medical image analysis.

Another potential area of research is the extension of our study to other types of cancer detection. Although our study focused specifically on oral squamous cell carcinoma, similar techniques could be employed for detecting other types of cancer, such as breast or lung cancer. Researchers could use the insights gained from our study to develop more accurate and reliable AI models for a range of medical applications.

Furthermore, our study highlights the potential impact of AI in the field of medical image analysis. By improving the accuracy of AI models for cancer detection, missed diagnoses could be reduced and earlier detection provided for patients. With the continued advancement of AI technology, there may be further opportunities to enhance the accuracy and efficiency of medical image analysis, which could have significant benefits for patients and healthcare providers alike.

ACKNOWLEDGEMENTS

- The Scholar Launch team for providing this opportunity and guidance
- Ramin Ramezani for greatly helping this project and teaching AI/ML concepts
- Michael Powell for assistance and support of this project

REFERENCE

Chaunzwa, Tafadzwa L, et al. “Deep Learning Classification of Lung Cancer Histology Using CT Images.” Scientific Reports, U.S. National Library of Medicine, 9 Mar. 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7943565/.

O'Shea, Keiron, and Ryan Nash. “An Introduction to Convolutional Neural Networks.” ArXiv.org, 2 Dec. 2015, https://arxiv.org/abs/1511.08458.

Rahman, Tabassum Yesmin. “A Histopathological Image Repository of Normal Epithelium of Oral Cavity and Oral Squamous Cell Carcinoma.” Mendeley Data, Mendeley Data, 7 Nov. 2019, https://data.mendeley.com/datasets/ftmp4cvtmb/1.

Uzun Ozsahin, Dilber, et al. “The Systematic Review of Artificial Intelligence Applications in Breast Cancer Diagnosis.” Diagnostics (Basel, Switzerland), U.S. National Library of Medicine, 23 Dec. 2022, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9818874/#:~:text=AI%20is%20accurate%20in%20 detection,is%20used%20in%20medical%20imaging.

Heart Attack Prediction Using Machine Learning

By Natalie Miner

AUTHOR BIO

Natalie Miner is currently a senior at Mound Westonka High School. She is particularly interested in applied math, machine learning, and business. Natalier has been part of her school's DECA program since 9th grade, which is a business club aimed at career development. She is a Minnesota first place State Recipient and International Finalist in DECA. Additionally, she is the founder and president of Like A Girl, an empowerment-focused project for middle and high school girls to help them build confidence and leadership skills. Natalie is also her school's Student Body President and is involved in Soccer, Hockey, and Track.

ABSTRACT

Heart disease is one of the leading causes of mortality due to heart attacks and other complications. When a person has a heart attack, their heart muscle begins to die because of a lack of blood flow due to a blocked or reduced blood supply. The most severe effects of a heart attack can be reduced with prompt identification and treatment. Machine learning can help medical professionals make an earlier diagnosis of heart disease and start treatment, reducing complications and saving lives. Using various attributes relevant to detecting heart disease, a machine-learning model was created that predicts a person's risk of heart disease. The logistic regression model created in the study has an overall accuracy of 85%. Ultimately, this model could play an important role in predicting heart disease and preventing the severe effects of a heart attack.

INTRODUCTION AND BACKGROUND

A heart attack is a life-threatening medical emergency that requires immediate treatment. When a person has a heart attack, they experience a sudden blockage or reduction of the flow of oxygen-bearing blood to the heart. Heart attacks are often a direct result of heart disease. This occurs when coronary arteries that supply the heart muscle with blood flow become narrowed by a buildup of fat, cholesterol, and other substances that are called plaque (Mayo Clinic). The buildup of plaque, called Atherosclerosis, slowly progresses to narrow the artery, but if the plaque ruptures a blood clot can form on it suddenly, blocking all blood flow in that artery, and causing a heart attack. According to the Centers for Disease Prevention and Control, more than 800,000 people in the United States have a heart attack yearly. Furthermore, about every 40 seconds, someone in the United States has a heart attack. Factors including a person's age, gender, lifestyle habits, and other medical conditions can raise a person's risk of heart disease and heart attacks.

A heart needs oxygen to survive; if blood flow is not restored quickly, the heart muscle will begin to die. The amount of damage to the heart muscle depends on the time between the loss of blood flow and treatment. Therefore, prompt treatment is needed to prevent death or severe damage. Half the deaths from a heart attack occur in the first 3 or 4 hours after symptoms begin (Cedars Sinai), so receiving treatment quickly and efficiently for a heart attack is a crucial part of limiting severe damage to the heart and death. Machine learning can help position professionals to make an earlier diagnosis and prevent severe effects.

Machine learning is a data analysis method that focuses on building systems that are able to learn and adapt without following explicit instructions. It uses algorithms and statistical models to analyze and predict patterns in a set of data. In a healthcare setting, machine learning can help medical professionals make quicker, more accurate diagnoses leading to improved patient outcomes. The primary purpose of this study is to use machine learning to create an efficient logistic regression model that predicts a person's risk of a heart attack.

METHODS

Exploration of the dataset

The dataset is from the UCI Machine Learning Repository. The dataset includes 13 features, 7 categorical and 6 numerical, and one label. There is information from 303 different patients in the dataset and there are no missing values. Most of the features are normally distributed with no significant skewing as shown in Table 1 below. A small number of features did not distribute normally.

	count	mean	std	min	25%	50%	75%	max
age	303.0	54.366337	9.082101	29.0	47.5	55.0	61.0	77.0
sex	303.0	0.683168	0.466011	0.0	0.0	1.0	1.0	1.0
cp	303.0	0.966997	1.032052	0.0	0.0	1.0	2.0	3.0
trtbps	303.0	131.623762	17.538143	94.0	120.0	130.0	140.0	200.0
chol	303.0	246.264026	51.830751	126.0	211.0	240.0	274.5	564.0
fbs	303.0	0.148515	0.356198	0.0	0.0	0.0	0.0	1.0
restecg	303.0	0.528053	0.525860	0.0	0.0	1.0	1.0	2.0
thalachh	303.0	149.646865	22.905161	71.0	133.5	153.0	166.0	202.0
exng	303.0	0.326733	0.469794	0.0	0.0	0.0	1.0	1.0
oldpeak	303.0	1.039604	1.161075	0.0	0.0	0.8	1.6	6.2
slp	303.0	1.399340	0.616226	0.0	1.0	1.0	2.0	2.0
caa	303.0	0.729373	1.022606	0.0	0.0	0.0	1.0	4.0
thall	303.0	2.313531	0.612277	0.0	2.0	2.0	3.0	3.0
output	303.0	0.544554	0.498835	0.0	0.0	1.0	1.0	1.0

Table 1: Distribution of each feature in the dataset

The understanding and interpretation of the features in this dataset are crucial to interpreting patterns within the data and correlations between different features. Each feature is shown to have an important role in the development of heart disease.

1. *Age (age) - numerical*

As a person's age increases, their risk of heart attack also increases. Aging causes changes in the heart and blood vessels that may increase a person's risk of heart disease.

2. *Sex (sex) - categorical*

Heart attacks are twice as common in men than in women. Women's naturally occurring hormone levels may protect against heart disease until menopause. After menopause, a woman's risk increases to match that of men (National Institutes of Health). There is a strong correlation between heart disease and sex as shown in Figure 1, where the orange bars signify a high risk of heart disease. A value of 0 indicates female and a value of 1 indicates male.

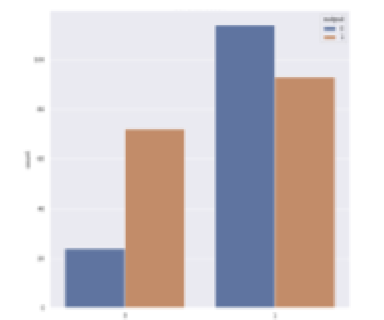


Figure 1: Correlation between heart disease risk and sex.

3. *Chest pain (cp) - categorical*

Most heart attacks include symptoms of chest pain or discomfort. This discomfort can feel like uncomfortable pressure, squeezing, fullness, or pain. This can be seen in four different forms: typical angina (0), atypical angina (1), non-anginal pain (2), or some may be asymptomatic (3). There is a correlation between chest pain and a risk of heart attack as shown in Figure 2. The orange bars signify a high risk of heart disease.

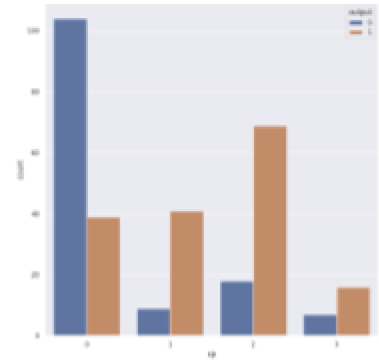


Figure 2: Correlation between heart disease risk and chest pain.

4. *Blood pressure (trtbps) - numerical*

High blood pressure can cause plaque to build up in the arteries. Therefore, the flow of blood through the heart muscle is interrupted, resulting in a heart attack (National Heart Association). A normal, healthy adult has an average blood pressure of approximately 120/80 mm Hg. Blood pressure above this can signify a risk of heart attack. The blood pressure is measured in millimeters of mercury (mm Hg).

5. *Cholesterol (chol) - numerical*

High levels of cholesterol can result in fatty deposits in the blood vessels. These deposits can make it difficult for enough blood to flow through the arteries. Cholesterol is measured in milligrams per deciliter and is fetched by a BMI sensor.

6. *Fasting blood sugar (fbs) - categorical*

High fasting blood sugar levels can lead to the buildup of plaque in the arteries and interrupt blood flow. A fasting blood sugar above 120 mg/dl is an indicator of diabetes and contributes to a higher risk of heart disease (National Institutes of Health). A value of 0 indicates blood sugar above 120 mg/dl and a value of 1 indicates a value lower.

7. *Resting electrocardiographic results (restecg) - categorical*

ECG is a test that records the electrical activity of the heart while a person is at rest. Abnormal tests can be used as evidence of coronary heart disease. A value of 0 indicates normal results, a value of 1 indicates ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV, and a value of 2 indicates probable or definite left ventricular hypertrophy by Estes' criteria. A correlation can be seen between these results and heart disease in Figure 3, where the orange bars signify a high risk of heart disease.

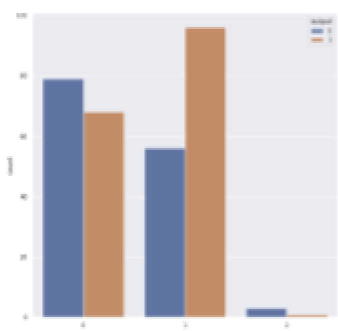


Figure 3: Correlation between heart disease risk and resting electrocardiographic results.

8. *Maximum heart rate achieved (thalach) - numerical*

The maximum heart rate achieved through exercise or stress can be an indicator of the heart's strength and ability to handle exertion.

9. *Exercise-induced angina (exng) - categorical*

Exercise-induced angina is chest pain that is a result of exercise or extreme stress. The heart muscle does not get enough blood or oxygen that it needs at a high activity level. A value of 0 indicates no and a value of 1 indicates yes. Figure 4 shows the correlation between exercise-induced angina and the risk of a heart attack. The orange bars signify a high risk of heart disease.

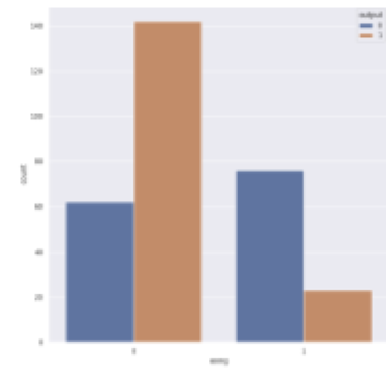


Figure 4: Correlation between heart disease risk and exercise-induced angina.

10. *ST depression induced by exercise relative to rest (oldpeak) - numerical*

ST depression occurs when the ST segment is abnormally low and appears below the baseline in a person's results. ST depression can indicate a lack of sufficient blood flow to the heart muscle (National Institutes of Health).

11. *The slope of the peak exercise ST segment (slp) - categorical*

The slope of the peak exercise ST segment can indicate the relative oxygen demand by the heart during exercise. This can reveal the overall health and condition of the heart. A value of 0 indicates upsloping, a value of 1 indicates a flat slope, and a value of 2 indicates downsloping.

12. *Number of major vessels colored by Fluoroscopy (caa) - numerical*

The number of major vessels colored by fluoroscopy reveals the measure of the presence of disease in the major blood vessels to the heart. The higher the number, the higher risk of severe disease.

13. *Thallium stress test (thall) - categorical*

A thallium stress test is an imaging test that indicates how well blood flows into the heart while exercising or at rest. This test can show areas of the heart muscle that aren't receiving enough blood, which is a sign of heart disease according to UCSF Health. A value of 1 indicates normal flow, a value of 2 indicates a fixed defect and a value of 3 indicates a reversible defect. A fixed defect poses a more significant problem. The orange bars signify a high risk of heart disease.

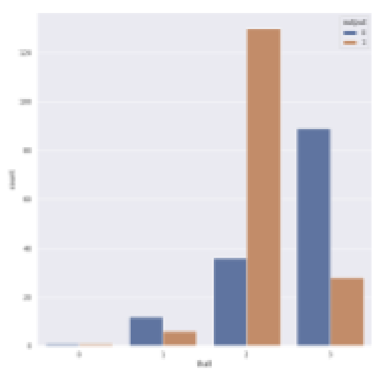


Figure 5: Correlation between heart disease risk and thallium stress test.

14. *Target (diagnosis of heart disease) - categorical*

This is the outcome of the prediction. A value of 0 indicates no significant heart disease and a value of 1 indicates a significant risk of heart disease.

PREPARATION OF DATASET FOR MACHINE LEARNING

Some features in this dataset are multicategorical. Because of this, the data must be one hot encoded so that it can be used in the model. One hot encoding will create new columns as much as the number of unique categories for the feature. The one hot encoded dataset can be seen in Table 2. This creates a form that can be provided to machine learning algorithms to run a prediction model.

	age	trtbps	chol	thalachh	oldpeak	caa	output	sex_1	cp_1	cp_2	cp_3	fbs_1	restecg_1	restecg_2	exng_1	slp_1
0	63	145	233	150	2.3	0	1	1	0	0	1	1	0	0	0	0
1	37	130	250	187	3.5	0	1	1	0	1	0	0	1	0	0	0
2	41	130	204	172	1.4	0	1	0	1	0	0	0	0	0	0	0
3	56	120	236	178	0.8	0	1	1	1	0	0	0	1	0	0	0
4	57	120	354	163	0.6	0	1	0	0	0	0	0	1	0	1	0

Table 2: One hot encoded dataset.

There are some significant numerical outliers in the data. These values can be deleted to create a more accurate model. Figure 6 shows the code below where these outliers are eliminated by calculating the upper and lower boundary and deleting values outside of this range.

```
for i in numeric_list:
    Q1 = np.percentile(df.loc[:, i],25)
    Q3 = np.percentile(df.loc[:, i],75)

    IQR = Q3 - Q1

    upper = np.where(df.loc[:, i] >= (Q3 + 2.5*IQR))
    lower = np.where(df.loc[:, i] <= (Q1 - 2.5*IQR))

    df.drop(upper[0], inplace = True)
    df.drop(lower[0], inplace = True)
```

Figure 6: Eliminate outliers

As a result, there are 298 data points without changing the number of features in the dataset.

MACHINE LEARNING MODEL

To start the prediction model, the data set was split into a training set and a validation set, with the training set being 75% and the validation set being 25%. The training set is used to train the model for the prediction of heart disease risk. The validation is then used to validate that the model accurately represents the data. This helps to check for overfitting later in the project, where the model gives accurate predictions for the training data, but not for any new data. The code for this step is seen in Figure 7.

Figure 7: Check that the dataset is balanced and create training and validation sets

The data was scaled using StandardScaler as seen in Figure 8. This step was important because it helps the machine learning model interpret features that have different ranges/magnitudes and interpret values on the same scale.

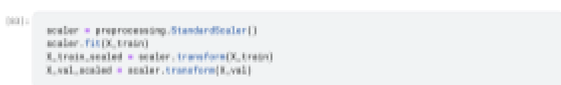


Figure 8: Scale the dataset

A logistic regression model was then created as seen in Figure 9. Logistic regression models are helpful in solving binary classification problems. The model can take into consideration multiple input criteria. An existing dataset is used to train a model to classify new data as either at high or low risk of heart disease. To do this, the model will estimate the probability of heart disease occurring (bounded between 0 and 1). For binary classification, a probability less than 0.5 will predict 0 (low heart disease risk) while a probability greater than 0.5 will predict 1 (high heart disease risk). Since there is a relatively small amount of data in the dataset, the model will be kept simple to avoid overfitting.

Figure 9: Logistic regression model

RESULTS

When looking at the results of the logistic regression model, it is important to compare the accuracy of the training and validation sets. We can check for overfitting using the validation and training errors seen in Figure 10 below. The model was able to avoid overfitting by confirming that the errors of the validation and training sets were not significantly different. We found that the validation error is approximately 0.379 and the training error is approximately 0.332. This shows a small but insignificant amount of overfitting.

Figure 10: Comparison of validation error and training error of logistic regression model

The logistic regression model created in this project to predict a person's risk of a heart attack is overall significantly accurate. We can look at the overall accuracy of the trained model using the classification report shown in Figure 11 below. This report provides important values such as accuracy, precision, and recall. The overall accuracy of this model was 0.85, indicating that it correctly classified 85% of the patients as either at a high or low risk of developing a heart attack. This accuracy score was relatively high given the size and dimensions of the dataset used in the project.

Figure 11: Classification report of logistic regression model

Importantly, by creating a prediction model that has a high performance level, this model has the potential to save lives by identifying individuals at a high risk of heart disease at an early stage. Early identification of at-risk individuals allows for timely interventions such as lifestyle modifications, medication management, and targeted medical interventions. Additionally, we can look at the features that are most impactful in the prediction of heart disease. Examples of these include age, blood pressure, cholesterol, and blood sugar. By intervening at an early stage of heart disease, medical professionals can reduce the number of heart attacks and their severe effects.

CONCLUSION

The machine learning model demonstrates the significant potential of machine learning algorithms to predict heart disease risk. By utilizing a dataset with relevant patient information, a model was created that could potentially save lives. Furthermore, each feature in the dataset correlates to heart disease risk, which was important to the machine learning model. Implementing a larger dataset to the logistic regression model created in this project could increase the overall accuracy of this model and help decrease overfitting. This study emphasizes the potential of early detection and intervention in preventing severe heart attacks using machine learning. This early detection may represent a useful tool to implement preventive measures in high-risk patients detected by this tool.

REFERENCE

Mayo Clinic Staff. (2022, May 25). Coronary Artery Disease. *Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613#:~:text=CAD%20happens%20when%20coronary%20arteries,or%20even%20a%20heart%20attack.*

Tsao, C. (2023, May 15). Heart Disease Facts. *Centers for Disease Prevention and Control. https://www.cdc.gov/heartdisease/facts.htm#:~:text=In%20the%20United%20States%2C%20someone,heart%20attack%20every%2040%20seconds.&text=Every%20year%2C%20about%20805%2C000%20people,States%20have%20a%20heart%20attack.&text=Of%20these%2C,are%20a%20first%20heart%20attack*

Cedars Sinai Staff. (2022). Heart Attack (Mycordial Infarction) *Cedars Sinai. https://www.cedars-sinai.org/health-library/diseases-and-conditions/h/heart-attack-myocardial-infarction.html#:~:text=Half%20the%20deaths%20from%20a,an%20ambulance%20with%20trained%20personnel.*

Harvard Health Staff. (2017, March 25). Gender Matters: Heart Disease Risk in Women. *Harvard Health Publishing.* https://www.health.harvard.edu/heart-health/gender-matters-heart-disease-risk-in-women

Ryczkowska, K. (2022, December 10). Menopause and Womens' Cardiovascular Health: Is It Really and Obvious Relationship? *National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10074318/*

American Heart Association Editorial Staff. (2022, December 2). What is a Heart Attack? *American Heart Association. https://www.heart.org/en/health-topics/heart-attack/about-heart-attacks*

Park, C. (2013, February 12). Fasting Glucose Level and Risk of Incident Atherosclerotic Cardiovascular Diseases. *National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/23404299/*

Boyette, L. (2022, July 11). Physiology, Myocardial Oxygen Demand. *National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK49989/*

UCSF Health Staff. (2018, June 18). Nuclear Stress Test. *UCSF Health. https://www.ucsfhealth.org/medical-tests/nuclear-stress-test*

Music therapy and Postnatal Depression: a Literature Review

By Xinyang Wang

AUTHOR BIO

Xinyang (Francesca) Wang is a 16-year-old junior at Guanghua Cambridge International School in Shanghai. She is passionate about psychology (clinical psychology in particular) and wishes to delve deeper in this field, and hopefully to be able to help a wider range of people in need in the future. In school, she is the founder and president of the French Club, sharing her self-taught French knowledge with club members. Meanwhile, she is a member of a non-profit educational organization called SoulShaper, which was founded by high schoolers and provides free online English courses to children all around China. In her free time, you can find her watching Formula 1 races and translating F1 news articles for Chinese fans.

ABSTRACT

Postnatal depression is a type of depression occurring after a woman has given birth. Music therapy, which means using music to improve health, is considered an alternative treatment for postnatal depression. This literature review aims to investigate whether and to what extent music therapy has an effect on reducing postnatal depression symptoms. While some results show that music therapy is potentially beneficial, the research done on this topic is very limited and future studies should be done to further explore the impact of music therapy on postnatal depression.

INTRODUCTION

Music has been showing its relationship with improved physical and mental health for many decades (Sanfilippo et al., 2021). In everyday life, people use music to adjust their emotions (Lin et al., 2019); in medical and other situations, music is further divided into several categories, including music therapy, community music, music medicine, and music education (MacDonald, 2013).

Music therapy, a type of therapy where a therapist uses music to help patients achieve certain health goals (American Music Therapy Association, 2005), has been shown to be helpful for treating many ailments. For instance, as mentioned in a mini-review by Kurdi & Gasti (2017), patients with Parkinson's disease show restored motor capacities after receiving music therapy. Also, music therapy has been found to be effective in helping autistic children to establish and improve their social and communicative skills (Kurdi & Gasti, 2017). Improved health conditions are seen in the elderly as well, after a month of music therapy exposure (Kurdi & Gasti, 2017).

Depression, or major depressive disorder, is one of the most prevalent mental disorders. Depression negatively influences how a person perceives the world and how they act (Torres, 2020). Postnatal depression is "a depressive state experienced by the mother within the first year following the delivery of a baby" (Terry, 2012). Mothers with postnatal depression are likely to experience a lower quality of sleep, a higher anxiety level, and a higher stress level (Stewart & Vigod, 2016). The mother-infant bonding may also be influenced, which will even affect the infant's future development (Netsi et al., 2018).

Treatments for postnatal depression are broadly divided into four categories: pharmacological, psychological, psychosocial, and complementary and alternative medicines (Terry, 2012). While pharmacological and psychotherapy and psychosocial therapies are already being relatively widely used, complementary and alternative medicines, including music therapy, are being researched and put into use more often, as they are less expensive, easier to access and have minor side effects (Lin et al., 2019; Terry, 2012). The objective of this literature review is to explore whether using music therapy can be successful in treating postnatal depression, and if so, to what extent can it be effective for reducing postnatal depression symptoms.

METHODOLOGY

When conducting this literature review, searching engines and databases such as Google Scholar, PubMed, and Refseek were used. Keywords used when searching for sources include music therapy, postnatal depression, postpartum depression, music, depression, treatments for postnatal depression, music therapy and postnatal depression, music therapy and depression. As this literature review focuses only on the effect of music therapy on postnatal depression during the postpartum period, sources analyzing the effect of the use of music interventions before or during pregnancy were excluded. When discussing relevant literature which includes music interventions to postnatal depression, but does not necessarily outline the word "music therapy", the term "music therapy" was still used.

DISCUSSION

Music Therapy

Music Therapy is defined as a licensed music therapist using music interventions to help patients achieve certain health goals within "a therapeutic relationship" (American Music Therapy Association, 2005). However, there are also articles arguing that music therapy does not necessarily need to be carried out with a licensed music therapist, or within a "theoretical framework", as long as the aim of using music interventions is to improve patients' mental or physical health (Terry, 2012).

MacDonald (2013) points out in a review that there is also overlap between music therapy and other forms of music, which are community music, music medicine, music education, and everyday use of music. In particular, the overlap between music therapy and community music is mostly significant. Instead of a licensed music therapist, community music activities are usually conducted by a community leader who has the experience of music instrument playing (Sanfilippo et al., 2021). The biggest difference between music therapy and community music, according to MacDonald (2013), is their different aims. For music therapy, as mentioned earlier, the primary goal is to bring positive effects to the clients' health. In contrast, community music does not list "therapeutic effects" as their main objective; they focus more on giving more opportunities of engaging in music activities and practices to people in their local communities (MacDonald, 2013). However, in most cases, community music still brings positive therapeutic effects despite these not being their first concern (MacDonald, 2013). Under these circumstances, it is hard to define clearly whether the music intervention involved is music therapy or community music.

Furthermore, the presence of a licensed music therapist may not be mandatory, as argued in some articles: there might be cases when a "non-professional" who is not holding a music therapist license strictly follows the rules of music therapy, and it would not be appropriate to not classify this act as "music therapy" (Leubner & Hinterberger, 2017). On top of this, there are significant differences in terms of "uniform standards or eligibility requirements" for music therapy and music therapists between different countries (Leubner & Hinterberger, 2017).

In general, there are two types of music therapy: active music therapy, and passive or receptive music therapy. In active music therapy, people "re-create, improvise or compose music", which requires a music therapist to use professional skills to help them act out their feelings and thoughts through music (Terry, 2012). While in passive, or receptive music therapy, people only listen to the music, either it is self-selected music or prescribed music by a music therapist (Lin et al., 2019). To this extent, the definitions of music therapy and music medicine are again overlapping as they both aim to have a positive impact on the health of the patients; music medicine is defined as patients listening to music chosen by a healthcare professional in "medical contexts" (Lin et al., 2019; MacDonald, 2013). Music medicine per se is a relatively more concentrated area and there are not many researchers working in this field, and it only overlaps with music therapy but not with other forms of music (MacDonald, 2013). In terms of the use of music medicine, helping patients who are receiving medical operations to alleviate their pain, anxiety, and distress is one of the most representative examples (MacDonald, 2013). Music therapy and music medicine together are called "music interventions", and their relation with improved physical and mental health has been shown by multiple studies (Lin et al., 2019).

In this literature review, in order to take every potential effect of music therapy into account, the term "music therapy" is used when discussing relevant literature which (i), does not point out explicitly which type of music or music interventions were involved, due to the overlap between different types of music and the ambiguity of their definitions, or (ii), indicates an absence of a licensed music therapist.

Music Therapy and Depression

Depression, as mentioned earlier, can bring negative impacts on a person's everyday life. In recent decades, more research on using music as an intervention to help patients suffering from depression has been done (Leubner & Hinterberger, 2017). In a review article by Leubner & Hinterberger (2017), 28 studies are analyzed to investigate the use and effect of music interventions on depression, and 26 of them show a "significant reduction" in depressive symptoms in participants in the music group compared to those in either the control or the comparison group. This includes a study in which older adults with depression are assigned into two music therapy groups (one with participants learning music therapy techniques with regular home visits by a music therapist and the other with the same techniques with moderate music therapist intervention) and one control group (Hanser & Thompson, 1994). The results show significantly improved depressive symptoms in the two music conditions than the control group (Hanser & Thompson, 1994). Another study discussed by Leubner & Hinterberger (2017) compares the effect of music therapy on depression with psychotherapy, and there are fewer depressive symptoms in participants in the music therapy group compared to those in the psychotherapy group (Castillo-Pérez et al., 2010). Multiple studies discussing the effect of music therapy on depressive symptoms in patients with other forms of disorders or even people with no diagnosed disorder (such as psychiatric patients, patients with cancer, the elderly, and prisoners.) are also analyzed by Leubner & Hinterberger (2017), and most of them prove the statement that music therapy has a positive impact on depressive symptoms. From these studies, it can be determined that music therapy can be used as a potential treatment for depression. Further, it can be inferred that music therapy may also be used as a treatment for postnatal depression, a type of mental disorder that falls under the big category of depression.

Postnatal Depression

Postnatal depression, or postpartum depression, is a type of depression diagnosed mostly within 6 months after giving birth, and any diagnosis in the first year after birth can be classified as postnatal depression (Bell & Andersson, 2016). Different from what is called "baby blues", postnatal depression lasts for a longer time and is not only caused by sudden and significant hormonal changes during the perinatal period. Postnatal depression also differs from postnatal psychosis which is often linked to suicide and infanticide, as postnatal depression is less serious but more common (Bell & Andersson, 2016).

According to Terry (2012), there is no factor that is known to directly cause postnatal depression to occur, but there are some risk factors which can contribute to a higher risk of getting postnatal depression. Ghaedrahmati et al. (2017) divide the risk factors of postnatal depression into five main categories: psychological, biological, obstetric, social, and lifestyle.

Psychological risk factors can be further divided into two aspects; one is a past history of depression, and the other is the way in which the mother perceives her birth experience. Mothers who are previously diagnosed with depression, either before pregnancy or prenatally, are shown to be more easily influenced by the hormonal changes after birth and thus more likely to develop postnatal depression, instead of developing only mild and temporary "baby blues" or maintaining a normal mental state. The mother's attitude toward her birth experience can be affected by factors such as experience of sexual abuse, unwanted child gender, and confidence towards parenting.

Biological factors include the age at pregnancy, blood glucose level, and hormone levels which include serotonin and tryptophan, oxytocin, estrogen, and several others. Obstetric factors refer to the number of children a woman has given birth to, the mode of birth which indicates the risk of pregnancy, and a disparity between the mother's expectations and real birth events.

Postnatal depression is also related to the social support a mother receives, in terms of "emotional support, financial support, intelligence support, and empathy relations".
A healthy lifestyle can possibly reduce the chance of the development of postnatal depression, which takes a diet with all nutrients required, a sufficient quality and quantity of sleep, and regular exercise into account. A more detailed discussion of the risk factors of postnatal depression can be found in the narrative review by Ghaedrahmati et al. (2017).

Postnatal depression not only has negative impacts on the mother, i.e. sleep difficulties, a higher anxiety level, and other common depressive symptoms (Stewart & Vigod, 2016), it is also linked to poor mother-infant bonding and adverse child outcomes, including difficulties in understanding and cognition, or higher chances of depression (Netsi et al., 2018). Thus, the importance of finding effective and reliable treatments for postnatal depression goes without saying.

Treatments for Postnatal Depression

There are four types of treatments for postnatal depression: pharmacological, psychological, psychosocial therapies, and complementary or alternative medicines (Terry, 2012).

Taking antidepressants, which is an example of pharmacological treatment, is the most common treatment for postnatal depression. Antidepressants, i.e. medications which are used against depression, are often the first choice for mothers with moderate-severe symptoms of postnatal depression, although its effectiveness and safety have not yet been approved by enough research (Molyneaux et al., 2014; Molyneaux et al., 2018). Meanwhile, studies have shown that there are serious side effects brought by antidepressants, such as drowsiness and headaches, which can heavily influence a mother's normal activity; antidepressants can even endanger the infant's health as they may be excreted into breast milk (Terry, 2012).

Psychological treatment is a non-medical option that is chosen by many mothers with postnatal depression because it does not have any side effects neither to themselves nor to the infant (Terry, 2012). As suggested by Stephens et al. (2016), psychological therapies include "support groups, counseling, cognitive behavioral therapy, interpersonal therapy, and psychodynamic therapy". Aside from being an effective treatment for postnatal depression, psychological methods can also be used to prevent postnatal depression before it occurs (Stephens et al., 2016).

Psychosocial therapy is another non-medication treatment for postnatal depression. As a "lack of social support" is considered one of the most important risk factors of postnatal depression, psychosocial therapy provides the exact social treatment for postnatal depression (Terry, 2012). These therapies involve talking with a psychosocial therapist, as well as engaging in group activities with other mothers with the same mental disease (Terry, 2012).

Complementary or alternative medicines are a relatively new form of treatment for postnatal depression, and there are more clinical psychologists who are willing to use both traditional treatments (pharmacological, psychological, or psychosocial) and complementary or alternative medicines such as music therapy, yoga, and massage (Terry, 2012). However, whether these types of treatment are effective for reducing postnatal depression symptoms and to what extent they are effective has not been assessed by enough research. As a result, this literature review aims to investigate if music therapy can be used as a tool to treat patients with postnatal depression.

Music Therapy and Postnatal Depression

Although music therapy has been used as a form of treatment for health problems and mental disorders, including cancer, coronary heart disease, autism, depression, anxiety, and many others (Lin et al., 2019; Terry, 2012), it appears surprising that there is very limited literature discussing the use of music therapy on mothers with postnatal depression. After a decade, what Terry (2012) concluded in a systematic review is still found to be true: there are "two distinct groups of research" broadly, one is "research which examined the effectiveness of common treatments for those individuals with PND, yet did not use music as an intervention", and the other is research that "assessed the use of music as a treatment for patients with depression, anxiety and stress", but "were not geared specifically for women with PND". However, there are several studies analyzing the effect of singing, which is a technique of active music therapy, on reducing postnatal depression symptoms.

In Fancourt & Perkins (2018)'s study, a randomized controlled trial was carried out on 134 mothers with moderate-severe symptoms of postnatal depression. The trial had 3 arms: group singing workshops, group play workshops, and usual care, with the purpose of making a comparison between group singing intervention and other types of interventions to postnatal depression. Each arm had five groups with 8 to 12 mothers for each group, and the program lasted for 10 weeks. In the singing group, which was also the experimental group, mothers listened to songs sung by the leaders, learned and sang songs with their babies, as well as making their own songs to express their feelings toward being a new mother. For the mothers in the creative play (comparison) group, they took part in activities such as doing arts and crafts with their babies and playing games with them. The singing group and creative play group were led by the same professional workshop leaders, in order to avoid any extraneous variables occurring because of the inconsistency. For mothers who were in the control group, usual care for postnatal depression was received. In the original report of the trial, Fancourt & Perkins (2018) classified the group singing intervention they used as "a novel psychosocial intervention", as social support was provided by the interaction of the mothers with postnatal depression. However, as mothers in the singing group listened to, learned, sung, and created songs with the aim of improving their postnatal depression symptoms, this type of intervention should also be considered as an example of using music therapy as a treatment, as mentioned earlier (although a credentialed therapist was not involved). The results of the randomized controlled trial showed a "significantly faster decrease" in the postnatal depression symptoms of mothers who were in the singing group, although all three arms experienced a decrease in the symptoms. One important result was, the speed of recovery of mothers in the creative play group did not differ significantly from mothers who received usual care. This implies that, even though social support was provided by both the singing group and play group, the singing group had its own unique characteristic that could contribute to the faster recovery speed of postnatal depression, which is probably linked to music itself. This adds to the evidence that music therapy can be used to treat postnatal depression. One noticeable result found in the post hoc tests done was, the recovery speed of the singing group was not statistically significantly faster than the creative play group, suggesting that social support was still a factor to be considered when treating patients with postnatal depression, possibly due to the fact that a lack of social support can be one of the origins of the development of postnatal depression.

A comparative qualitative study was later conducted with 54 mothers who had completed the 10-week group singing program, in order to explain the characteristics of the group singing intervention which helped to accelerate the improvement of postnatal depression (Perkins et al., 2018). Semi-structured interviews with the mothers reviewed several characteristics of the group singing intervention, which involved giving mothers a sense of "achievement and identity", as well as the confidence and enhanced ability of being a mother, and other features such as giving mothers a chance to take a break from the role of a mother and to just be themselves.

Regarding limitations of this study, the sample consisted mainly of well-educated mothers (Fancourt & Perkins, 2018), which decreases the generalizability to all women with postnatal depression, as less-educated mothers were not included. Also, the relatively small sample size (134 mothers in the randomized controlled trial and 54 mothers in the qualitative study) is also a limitation and more studies need to be done using a larger sample size.

In another randomized controlled trial conducted by Wulff et al. (2021), the influence of maternal singing, which is also a type of active music therapy, on postnatal depression was investigated. However, the focus of this trial was not only on the postnatal depression symptoms, but also on maternal well-being and mother-infant bonding. Several questionnaires and scales including the Edinburgh Postnatal Depression Scale were sent out before the singing intervention, and after the intervention, which was 10 weeks later. Although there was no significant difference in the depressive symptoms reported by the mothers in the singing group and control group, the correlation analysis suggested fewer depressive symptoms when there is a higher frequency of mothers singing to themselves. However, although the depressive symptoms were assessed using the Edinburgh Postnatal Depression Scale, as the mothers recruited in this study were not all diagnosed with postnatal depression, the results of this trial could have been affected by this limitation.

CONCLUSION

In all, postnatal depression is a type of depression affecting approximately 10% of the women in the world, which can be caused by several risk factors including psychological factors, biological factors, obstetric factors, social factors, and lifestyle. Postnatal depression can have a negative effect on the mother, as well as her children. Treatments available for postnatal depression are divided into four main categories: pharmacological, psychological, psychosocial, and complementary and alternative medicine. A treatment for postnatal depression which falls under the category of complementary and alternative medicine is called music therapy. Although commonly being defined as a licensed music therapist using music to help clients achieve specific personal health goals, the definition of music therapy is still disputed, mostly about whether a therapeutic relationship with a licensed music therapist is necessary or not. There are two types of music therapy, namely active and passive. In terms of these two types, the overlap between music therapy and other forms of music becomes more significant.

Despite that music therapy has been proven to have a positive effect on depression in general, and it can be extrapolated from this that music therapy may also be useful in treating postnatal depression, literature analyzing the use of music therapy on postnatal depression is very limited. The few existing research on this specific topic, although confirming that music therapy has the potential of helping to reduce postnatal depression symptoms, has several limitations, with regard to the relatively small-sized and unrepresentative sample, and the techniques of music therapy which were being used. More research should be done on this topic, to further prove the effectiveness of music therapy on postnatal depression and explore the extent to which it can be effective. Generally, current research and studies are only using group singing as a technique, which is a type of active music therapy, as a treatment for postnatal depression, while other techniques of active music therapy, and passive music therapy should be used in future studies. In addition, almost no current study on this topic highlights the use of a licensed music therapist, which suggests that the importance of the involvement of a music therapist can also be a research direction, considering the general definition of music therapy.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Roger Worthington and Dr. Bridget Hamill, for being her mentors on this research. They gave consistent help and advice throughout the research and construction of this paper, and the author sincerely appreciates this.

REFERENCE

American Music Therapy Association. (2005). What is Music Therapy | What is Music Therapy? | American Music Therapy Association (AMTA). Musictherapy.org. https://www.musictherapy.org/about/musictherapy/

Bell, A. F., & Andersson, E. (2016). The birth experience and women's postnatal depression: A systematic review. Midwifery, 39, 112–123. https://doi.org/10.1016/j.midw.2016.04.014

Castillo-Pérez, S., Gómez-Pérez, V., Velasco, M. C., Pérez-Campos, E., & Mayoral, M.-A. (2010). Effects of music therapy on depression compared with psychotherapy. The Arts in Psychotherapy, 37(5), 387–390. https://doi.org/10.1016/j.aip.2010.07.001

Fancourt, D., & Perkins, R. (2018). Effect of singing interventions on symptoms of postnatal depression: three-arm randomised controlled trial. The British Journal of Psychiatry, 212(2), 119–121. https://doi.org/10.1192/bjp.2017.29

Ghaedrahmati, M., Kazemi, A., Kheirabadi, G., Ebrahimi, A., & Bahrami, M. (2017). Postpartum depression risk factors: A narrative review. Journal of Education and Health Promotion, 6(60), 60. https://doi.org/10.4103/jehp.jehp_9_16

Hanser, S. B., & Thompson, L. W. (1994). Effects of a Music Therapy Strategy on Depressed Older Adults. Journal of Gerontology, 49(6), P265–P269. https://doi.org/10.1093/geronj/49.6.p265

Kurdi, M. S., & Gasti, V. D. (2017). Music and Health. Journal of Applied and Advanced Research, 2(2), 95. https://doi.org/10.21839/jaar.2017.v2i2.62

Leubner, D., & Hinterberger, T. (2017). Reviewing the Effectiveness of Music Interventions in Treating Depression. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.01109

Lin, C.-J., Chang, Y.-C., Chang, Y.-H., Hsiao, Y.-H., Lin, H.-H., Liu, S.-J., Chao, C.-A., Wang, H., & Yeh, T.-L. (2019). Music Interventions for Anxiety in Pregnant Women: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Journal of Clinical Medicine, 8(11), 1884. https://doi.org/10.3390/jcm8111884

MacDonald, R. A. R. (2013). Music, health, and well-being: A review. International Journal of Qualitative Studies on Health and Well-Being, 8(1), 20635. https://doi.org/10.3402/qhw.v8i0.20635

Molyneaux, E., Howard, L. M., McGeown, H. R., Karia, A. M., & Trevillion, K. (2014). Antidepressant treatment for postnatal depression. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.cd002018.pub2

Molyneaux, E., Telesia, L. A., Henshaw, C., Boath, E., Bradley, E., & Howard, L. M. (2018). Antidepressants for preventing postnatal depression. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.cd004363.pub3

Netsi, E., Pearson, R. M., Murray, L., Cooper, P., Craske, M. G., & Stein, A. (2018). Association of Persistent and Severe Postnatal Depression With Child Outcomes. JAMA Psychiatry, 75(3), 247. https://doi.org/10.1001/jamapsychiatry.2017.4363

Perkins, R., Yorke, S., & Fancourt, D. (2018). How group singing facilitates recovery from the symptoms of postnatal depression: a comparative qualitative study. BMC Psychology, 6(1). https://doi.org/10.1186/s40359-018-0253-0

Sanfilippo, K. R. M., Stewart, L., & Glover, V. (2021). How music may support perinatal mental health: an overview. Archives of Women's Mental Health, 24(5), 831–839. https://doi.org/10.1007/s00737-021-01178-5

Stephens, S., Ford, E., Paudyal, P., & Smith, H. (2016). Effectiveness of Psychological Interventions for Postnatal Depression in Primary Care: A Meta-Analysis. The Annals of Family Medicine, 14(5), 463–472. https://doi.org/10.1370/afm.1967

Stewart, D. E., & Vigod, S. (2016). Postpartum Depression. New England Journal of Medicine, 375(22), 2177–2186. https://doi.org/10.1056/nejmcp1607649

Terry, D. (2012). Singing the Blues: A Literature Review of the Effects of Music on Postnatal Depression.

Torres, F. (2020, October). What Is Depression? Psychiatry; American Psychiatric Association. https://www.psychiatry.org/patients-families/depression/what-is-depression

Wulff, V., Hepp, P., Wolf, O. T., Fehm, T., & Schaal, N. K. (2021). The influence of maternal singing on well-being, postpartum depression and bonding – a randomised, controlled trial. BMC Pregnancy and Childbirth, 21(1). https://doi.org/10.1186/s12884-021-03933-z

Prediction of Heart Failure Using Random Forest and XG Boost

By Aidan Gao

AUTHOR BIO

Aidan Gao is a sophomore attending The Westminster Schools in Atlanta, Georgia. He is interested in the uses of computer science and machine learning in the medical realm. He hopes to study computational medicine in the future.

ABSTRACT

Heart Failure (HF), a type of Cardiovascular Disease (CVD), is a prevalent illness that can lead to hazardous situations. Each year, approximately 17.9 million patients globally die of this disease. It is challenging for heart specialists and surgeons to predict heart failure accurately and on time. Fortunately, there are classification and prediction models available that can assist the medical field in efficiently using medical data. The objective of this study is to enhance the accuracy of heart failure prediction by prediction modeling a Kaggle dataset composed of five sets of data over 11 patient attributes. Multiple machine learning approaches were utilized to understand the data and forecast the likelihood of heart failure in a medical database. The results and comparisons show a definite increase in the accuracy score of predicting heart failure. Integrating this model into medical systems would prove beneficial for aiding doctors predictions of heart disease in patients

Keywords: machine learning, heart failure, diagnosis, prediction modeling, binary classification, random forest, XGBoost, cardiovascular disease

INTRODUCTION

The primary cause of heart stroke is the obstruction of arteries, also known as cardiovascular disease or arterial hypertension World Health Organization (n.d.). Heart disease affects approximately 26 million people worldwide, and this number is expected to rise rapidly if effective measures are not taken (Savarese & Lund, 2017). Unhealthy food, tobacco, excessive sugar, and obesity are common contributors to heart disease (Benjamin et al., 2019). Pain in the arms and chest are common symptoms, but the disease often presents with different symptoms based on sex and age. In addition to maintaining a healthy lifestyle and diet, timely diagnosis and comprehensive analysis are critical factors in identifying heart disease. However, many patients undergo multiple tests that can be physically and financially burdensome. Proper analysis of this type of data can improve the diagnosis process and assist heart surgeons. Previous research has used various techniques such as Random Forest, Support Vector Machine, and other AI classification models (Alotaibi, 2019). This study aims to surpass previous studies' random forest model accuracy in order to better predict heart failure before it manifests.

LITERATURE REVIEW

Previous work has utilized a subset of the dataset used in this paper to predict heart failure. The University of California Irvine (UCI) used Decision Tree, Logistic Regression, Random Forest, Naïve Bayes, and SVM reaching results around the ~85% accuracy mark. Through the use of tenfold cross validation and an enlarged dataset, previous studies have enhanced the accuracy of the UCI models (see Table. 1) (Alotaibi, 2019).

Table 1. Performance Comparison

Model	Alotaibi (Alotaibi, 2019)	UCI, 2019 (Bashir et al., 2019)	UCI, 2017 (Ekiz and Erdogmus, 2017)	UCI, 2017 (Ekiz and Erdogmus, 2017)
Decision Tree	93.19%	82.22%	60.9%	67.7%
Logistic Regression	87.36%	82.56%	65.3%	67.3%
Random Forest	89.14%	84.17%	X	X
Naïve Bayes	87.27%	84.24%	X	X
SVM	92.30%	84.85%	67%	63.9%

DATA OVERVIEW

The dataset utilized in this paper is collected from Kaggle under the name "Heart Failure Prediction Dataset" (Ortega, 2021). The dataset combines five datasets with over 11 common attributes. These five datasets combine data from Cleveland, Hungarian, Switzerland, Long Beach VA, and Stalog datasets. In total, the dataset contains 918 rows. Row definitions are provided below:

1. Age: age of the patient [years]
2. Sex: sex of the patient [M: Male, F: Female]
3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
4. RestingBP: resting blood pressure [mm Hg]
5. Cholesterol: serum cholesterol [mm/dl]
6. FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
7. RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or

depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]

8. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
9. ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
10. Oldpeak: oldpeak = ST [Numeric value measured in depression]
11. ST_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
12. HeartDisease: output class [1: heart disease, 0: Normal]

METHODS

Data Preprocessing

Before putting the data into the model, data preprocessing methods are applied in order to make it useful for modeling (Al-Mudimigh et al., 2009). Binary values, such as sex, were converted to binary numbers (1 and 0). ExerciseAngina was converted to binary as well. One hot encoding was employed in ChestPainType. The ordinal encoder from SKLearn was employed in the ordinal variables, ST_Slope and RestingECG.

This research focuses on two models to predict heart disease: Random Forest and XG Boost. Random Forest was implemented with the goal to surpass previous research accuracy with its cross-validation score. XGBoost was also implemented as a popular model among many Kaggle dataset winners. Both models were cross-validated ten times across an 80-20 split of training and test data, respectively. In order to further increase accuracy, HyperOPT (Komer et al., 2019) and RandomizedSearch (Agrawal, 2021) were used to finetune the hyperparameters for XGBoost and Random Forest, respectively.

Random Forest

The Random Forest algorithm is utilized to address classification issues. Its approach is based on ensemble learning, which combines multiple classifiers to enhance the algorithm's performance. The algorithm is composed of several Decision Trees classifiers to create a forest (Donges, 2018), each working on a subset of data, and the average is calculated to improve prediction accuracy. Rather than relying on the prediction of a single tree, the Random Forest algorithm combines the trees using an estimated outcome and voting procedure (Bashar et al., 2019). The model then considers predictions from each tree and determines the outcome based on majority voting.

XGBoost

XGBoost, which stands for "Extreme Gradient Boosting," is a popular machine learning model that has been used for a variety of tasks such as regression, classification, and ranking. XGBoost creates a model in the form of boosting an ensemble of weak classification trees by gradient descent which provides optimization to the loss factor (Cui et al., 2017). It is an ensemble learning algorithm that combines multiple decision tree models to improve the accuracy and robustness of predictions. XGBoost has gained popularity due to its speed, scalability, and performance. It uses a gradient-boosting framework and can handle missing values, regularization, and parallel processing. Additionally, it has various hyperparameters that can be tuned to achieve better performance. Overall, XGBoost is a powerful machine learning model that is widely used in industry and academia.

RESULTS

When looking at binary classification problems, there are four relevant metrics: true positives, true negatives, false positives, and false negatives. Out of these four metrics, the most harmful to prediction would be false negatives, or results that report no risk of heart failure despite the patient being at risk. The goal of this research was to reduce the number of false negatives and false positives in order to improve accuracy in predicting heart disease. Using these models and methods, the result was 91.56% accuracy after cross validation for XGBoost, with 9 false negative cases out of 181 cases (see Fig. 1). In Random Forest, there was 92.90% accuracy after cross validation with 6 false negatives out of 181 cases (see Fig. 2). In comparison to Alotaibi (2019), Random Forest performed considerably better, increasing from 89.14 to 92.90% accuracy. The accuracy increase may be due to the hyperparameter tuning for Random Forest or the cross-validation method. Also taken into consideration is the artificial addition of rows. In Alotaibi (2019), the size of the Cleveland data was too low to implement machine learning approaches. Alotaibi increased the size of the data artificially by randomizing values between minimums and maximums. The issue with randomizing these values is that there is no way to tell whether the target value is right for the artificial patient, thus creating noisy data that is useless for the model. This causes much of the data to carry either a random target value or no target value, which would cause the accuracy of the model to go down[1].

Another metric used in binary classification problems is the ROC curve, or the area under the receiver operating characteristic, a common metric for evaluating binary classification models. A model with a higher AUC is thought of as a better model (Javeed et al., 2019). This value was 0.92 for XGBoost and Random Forest. Both models performed well in terms of accuracy and the area under the ROC curve, indicating their effectiveness in predicting heart disease.

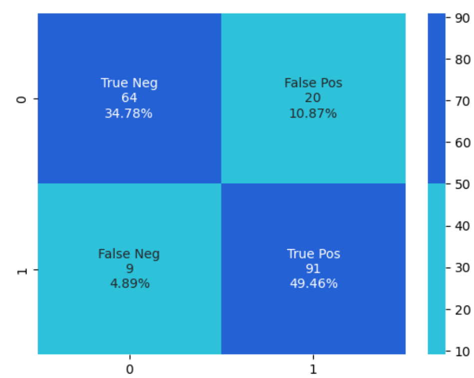


Fig. 1. Confusion Matrix of XGBoost Model

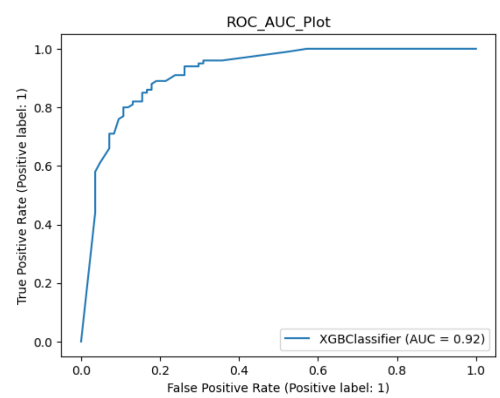


Fig. 2. ROC_AUC Plot of XGBoost

[1] This may not be the case, but there is no evidence in the paper or the references to explain the choice made by Alotaibi (2019).

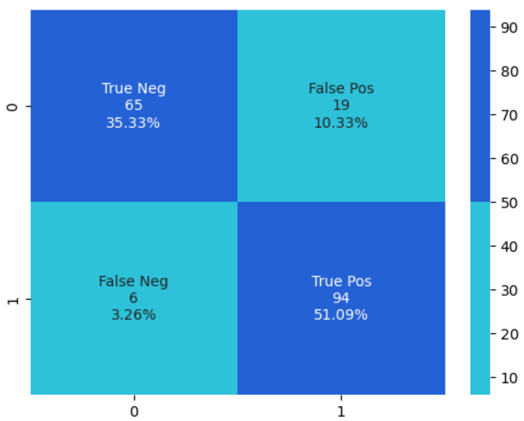


Fig. 3. Confusion Matrix of Random Forest Model

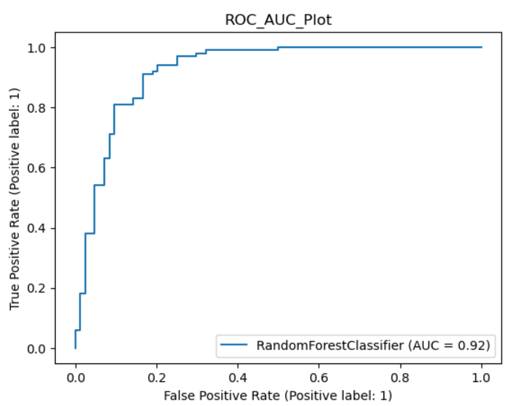


Fig. 4. ROC_AUC Plot of Random Forest Model

Table 2. Performance Comparison

Model	This Study	Alotaibi (Alotaibi, 2019)	UCI, 2019 (Bashir et al., 2019)	UCI, 2017 (Ekiz and Erdogmus, 2017)	UCI, 2017 (Ekiz and Erdogmus, 2017)
Decision Tree	X	93.19%	82.22%	60.9%	67.7%
Logistic Regression	X	87.36%	82.56%	65.3%	67.3%
Random Forest	92.90 %	89.14%	84.17%	X	X
Naïve Bayes	X	87.27%	84.24%	X	X
SVM	X	92.30%	84.85%	67%	63.9%
XGBoost	91.56 %	X	X	X	X

DISCUSSION

The results of this research reveal the potential use of machine learning in the landscape of cardiovascular healthcare. The degree of accuracy achieved by the current models opens up a pathway for research that may dramatically impact the diagnosis of heart disease. If integrated into medical information systems, these models could facilitate the collection and analysis of live data from patients, allowing doctors to make predictions in real-time. Hypothetically, machine learning models could form a network. Accessible to healthcare providers across the world, such a system could flag patients the models decide show significant risk of heart disease, prompting immediate intervention and follow-up. This would have a profound effect on the field of heart disease management, further shifting it from a reactive to a proactive field. Moreover, the system could be used in more rural, underserviced, areas where traditional diagnostic resources are unavailable.

This study helps to revolutionize early detection methods for heart disease, which in turn, significantly impacts patient outcomes. When applied on a larger scale, this approach could reduce mortality rates and enhance quality of life in patients diagnosed with this condition.

LIMITATIONS AND FURTHER WORK

These models are trained on historical data and their predictions are based on patient data patterns exhibited in past cases. However, upon application on a larger scale, much more patient data could be collected, training the model to keep up with modern times. One limitation of this study is that the dataset utilized to model was relatively small. Therefore, the dataset may not be able to accurately mirror large populations, a significant limitation for a model meant to be used at the population level. To increase the accuracy and reliability of these models, future research should aim to expand the dataset in not just rows but also columns. With the addition of more lifestyle variable columns, the model's accuracy is likely to rise. Future studies should explore handling outlier cases that do not align with patterns in historical data. This could involve using hybrid models that incorporate machine learning with other predictive tools. With further development, these models could potentially bring the rise of a new era of predictive healthcare in cardiology.

CONCLUSION

Heart failure is a significant health issue that affects millions of people worldwide. Machine learning models, specifically Random Forest and XGBoost, can be used to accurately predict heart failure based on medical data. These models can be integrated with medical information systems to improve the accuracy of predictions and assist healthcare providers in detecting heart disease early.

ACKNOWLEDGEMENTS

I would like to acknowledge Ms. Avi-Yona Israel for her amazing support and guidance in the writing process of this paper

REFERENCE

World Health Organization. (n.d.). Cardiovascular diseases. Retrieved March 15, 2023, from https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1

Savarese, G., & Lund, L. (2017). Global public health burden of heart failure. Cardiac Failure Review, 3(1), 7-11.

Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... & Virani, S. S. (2019). Heart disease and stroke statistics-2019 update: A report from the American Heart Association. Circulation, 139(10), e56-e528.

Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. International Journal of Advanced Computer Science and Applications, 10(6), 261-268. doi: 10.14569/IJACSA.2019.0100637

Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019). Improving heart disease prediction using feature selection approaches. In 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 619-623).

Ekiz, S., & Erdogmus, P. (2017). Comparative study of heart disease classification. In 2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting, EBBT 2017 (pp. 1-4).

Ortega, F. (2021). Heart failure prediction dataset. Kaggle. Retrieved February 25, 2023, from https://www.kaggle.com/fedesoriano/heart-failure-prediction.

Al-Mudimigh, F., Ullah, Z. S. A., & Saleem, A. (2009). A framework of an automated data mining systems using ERP model. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 1(5), 598-605.

Komer, B., Bergstra, J., & Eliasmith, C. (2019). Hyperopt-sklearn. In Automated Machine Learning: Methods, Systems, Challenges (pp. 97-111). Springer.

Agrawal, T. (2021). Hyperparameter optimization using scikit-learn. In Hyperparameter Optimization in Machine Learning (pp. 21-41). Apress.

Donges, N. (2018). The random forest algorithm. Towards Data Science. Retrieved March 15, 2023, from https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

Bashar, S. S., Miah, M. S., Karim, A. H. M. Z., Al Mahmud, A., & Hasan, Z. (2019). A machine learning approach for heart rate estimation from PPG signal using random forest regression algorithm. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-5).

Cui, Y., Cai, M., & Stanley, H. E. (2017). Comparative analysis and classification of cassette exons and constitutive exons. BioMed Research International, 2017, 1-8.

Javeed, A., Zhou, S., Yongjian, L., Qasim, I., Noor, A., & Nour, R. (2019). An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection. IEEE Access, 7, 180235-180243. doi: 10.1109/ACCESS.2019.2950861