

# Forecasting Stock Prices by Analyzing Announcements: The Case of Colombo Stock Exchange

Gayana Wickramarathna\*, <sup>2</sup>Hashan Ratnayake\*\*

\*Undergraduate Student, Rajarata University of Sri Lanka.

\*\*Lecturer, Rajarata University of Sri Lanka.

DOI: 10.29322/IJSRP.10.08.2020.p10450

<http://dx.doi.org/10.29322/IJSRP.10.08.2020.p10450>

**Abstract** -Buying and selling shares has become one of the most popular and lucrative investment decisions for investors. Unlike before, today's stock price decision has become a more complex task. Profitability of the stock market investment depends on the investor's decisions and is based on a mix of dynamic environmental factors. Stock price trends are repeatedly forecasted to extract useful patterns and predict their movements. Moreover, the stock market forecast always has some appeal. There are different approaches to stock price prediction and different forecasting methods are used by stock market analysts. As a replica of many scientific endeavors, several methods have been found to accurately predict stock prices. Most researchers have been used technical analysis to get more accurate results, while limited researchers used fundamental analysis. From Sri Lanka's perspective, there is no evidence on predicting stock prices using machine learning. The main objective of this paper is to measure the effect of the announcement on the price of the stock and fill in the gaps in the literature. The population of the current study is all listed companies on the Colombo Stock Exchange. Among them, Ceylon Tobacco Company PLC, Dialog Axiata PLC, and John Keells Holdings PLC were selected as the sample for this study by employing simple random sampling method and market capitalization. The analysis was conducted by employing natural language processing mechanisms and Random Trees classifier presented the best results. The findings facilitate to predict the share price of the Colombo Stock Exchange using machine learning techniques. This model was able to predict the stock price with 65% accuracy and would benefit all individual investors in the local stock market.

**Key Words**- Colombo Stock Exchange, Fundamental Analysis, Machine Learning, Random Trees, Stock Price Prediction

## I. INTRODUCTION

Many investors and brokers today use intelligent trading systems that help them predict stock prices based on various conditions, thereby helping them make direct investment decisions. An intelligent trader will predict the stock price and buy a stock before the price rises, or sell it before its value falls. Most investors have basic financial literacy but do not believe that they have sufficient basic knowledge about investing in the stock market [1]. Therefore, the conflict between knowledge and their trust leads them to misrepresent investments and rely on rumors about their investment decision. Lack of adequate knowledge of the capital market leads not only to loss but also to unattractive investment. Basic financial literacy is not an essential prerequisite for a good investment decision in the capital market, but it helps to make an investment decision and prevents a large loss [1].

The answer to this is the analysis of the financial statements. There are two types of stock analysis done by investors before investing in the stock; the first is the Fundamental Analysis, in this analysis investor's look at the intrinsic value of stocks, and performance of the industry, economy, political environment, etc. On the other hand, technical analysis is the analysis of stock prices by studying the statistics generated by past share prices and volume [2].

The Colombo Stock Exchange (CSE) is the leading stock exchange in Sri Lanka that offers investors buy and sell shares. It is one of the major exchanges providing an e-commerce platform in South Asia [2]. The official website of the Colombo Stock Exchange publishes announcements every day. They affect the share price of the company. This article analyzes the cse.lk announcement and predicts the impact on pricing. In recent years, the increasing importance of machine learning in various industries has enabled many vendors to apply machine-learning techniques in the field, some of which have yielded good results. Many researchers used technical analysis to learn machine learning. Limited researchers have conducted fundamental analyzes. There is a paucity of research on stock price forecasting using machine learning in the context of Sri Lanka.

## II. LITERATURE REVIEW

The Colombo Stock Exchange is the backbone of the Sri Lankan economy. It was founded in 1896 under the Share Brokers Association (SBA). Establishing a Formal Stock Exchange with the merger of the Colombo Stock Exchange (CSE) in 1985, the stock market was taken over by the Colombo Share Brokers Association. It currently has a membership of 15 companies and is licensed to operate as stockbrokers. In 1990, the business was renamed the Colombo Stock Exchange. The CSE introduced the Central Depository System and the clearance was automated. The Colombo Stock Exchange headquarters was opened at WTC in 1995. The Milanka Price Index was introduced in 1999 and opened several branches in the country. Regional branches contain Kandy, Jaffna, Anuradhapura, Kurunegala, Matara, Ratnapura, and Ambalantota [3]. The Colombo Stock Exchange (CSE) has 289 companies representing 21 GICS industry groups as of January 20, 2020, with a market capitalization of Rs. 2,748.10 Bn.

During the '50s, the use of probability theory and statistics for asset pricing models began to be applied by financial economics. These developments led to the discovery of the Capital Asset Pricing Model (CAPM) [4][5]. Initially rejecting CAPM and other statistical approaches, the Efficient Market Hypothesis (EMH) [6] argues that stock price movements are a function of rational expectations based on publicly known information from companies, and these expectations are almost immediately reflected in stock prices, for example in the price history. This means that there is no justification for modeling stock price changes using the price history when these changes have already been accommodated to stock prices.

In response to EMH, Behavioral Economics (BE) theories [7] argue that markets are ineffective and that the element of random walking can actually be explained by human behavior. Like, they make irrational and systematic mistakes. For EMH and BE reconciliation, the adaptive market hypothesis (AMH) Andrew [8] states that traditional models can be correlated with behavioral models, which implies that market efficiency levels are associated with environmental factors. The adaptability of market competitors, size of available profit opportunities and market participants.

Stock market exchange news are special messages containing mainly economic and political information. Some of them carry information that is important for market forecasting. There is a variety of financial information sources available on electronic versions of their daily problems. All of these sources include global and regional political and economic news, quotes from influential bankers and political politicians, and recommendations from financial analysts.

Chan, Chui, & Kwok [9] confirm the response to news articles. They have shown that economic news always has a positive or negative impact on the number of shares traded. They used important political and economic news as a proxy for public information. They found that news affects both trading activity, including return volatility, price volatility, number of traded stocks and frequency. Yu [10] present a text-mining-based framework for determining the sentiment of news articles and their impact on energy demand. The news sentiment is presented quantitatively and in a time series, comparing energy demand and price fluctuations.

In their research, Nagar & Hahsler [11] proposed an automated text-based approach to gathering news from various sources and forming a news corporation. Corpus is filtered into relevant sentences and analyzed using natural language processing (NLP) techniques. News Sentiment, a psychometric metric that uses positive and negative polarity terms, is proposed as a measure of the overall news corporation's mood. This research paper Shynkevich [12] studies how the results of financial forecasting can be improved when multiple levels of relevant news articles are used simultaneously for the target stock. They used multi-kernel learning methods to segment information extracted from five different news categories based on sectors, sub-sectors, industries, and so on. The newsletters are divided into five sections related to the target segment, its sub-industry, industry, group industry, and sector, and use a separate kernel to analyze them. Experimental results show that the simultaneous use of five research groups improves predictive performance compared to methods based on low news categories. The findings show that trade-offs yielded the highest predictive accuracy and trade-off when the five news categories were used with two separate kernels of polynomial and Gaussian types used for each news category.

Machine learning methods for stock price predictions are becoming popular. Machine learning is the scientific study of algorithm and statistical models which use computer systems to perform a particular task. Without using specific instructions [13]. Through various experiments, it is possible to test machine learning techniques and select the most suitable one for predicting stock prices [14]. An increasing trend is machine learning and the use of artificial intelligence to predict stock prices [15]. More and more researchers are investing their time each day to come up with strategies that can improve the accuracy of the stock forecasting model. Stock market forecasting has long been an interesting topic for researchers in various fields. In particular, a variety of studies is conducted to predict the stock exchange movement using machine learning algorithms like support vector machines (SVM) and reinforcement learning [2].

Developed the WEKA (Waikato Environment for Knowledge Analysis) at the University of Waikato, New Zealand; Written in Java and distributed under the terms of the GNU General Public License [16].

WEKA is tried and tested using a graphical user interface, standard peripheral applications, or open-source machine learning software accessible through the Java API. It is mostly used for teaching, research, and industrial applications and has a wide range of tools for standard machine learning tasks. Famous toolboxes like Sci-kit-learn, R, and Deeplearning4j. [17].

Singhal & Jena (2013) have explained that the basic concept is used to analyze, categorize and ultimately summarize data from different angles. They introduce the core principle of data preprocessing, classification, clustering and the introduction of the WEKA tool. WEKA is a data-mining tool. How to use the WEKA tool for these technologies is outlined? It provides the ability to categorize data across different algorithms.

In the fundamental analysis, Joshi[19] used Naive Bayes, Random Forest, and SVM and Random Forest show ability to predict the stock price more accurately than the other functions. Kalmegh[16] used REPTree, Simple Cart and Random Tree Classifiers to analyze the data sets. As a result, it is found that Random Tree algorithm performs best in categorizing all the News.

### III. Research Methodology

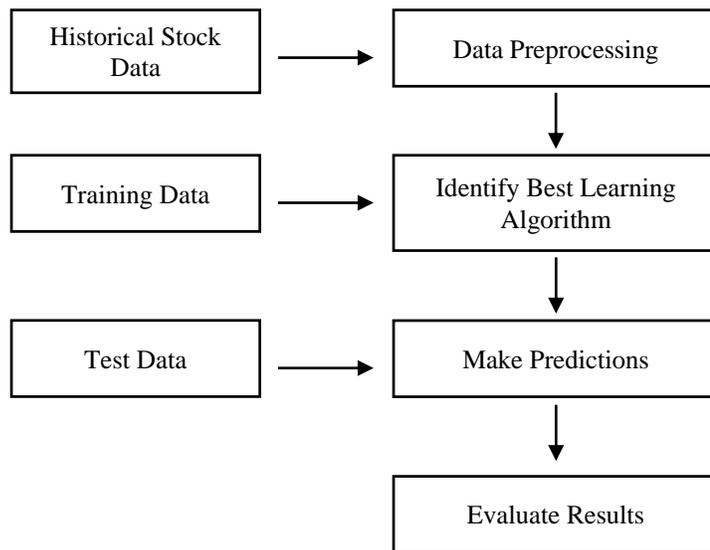


Figure 1: Framework

#### Step 01: Historical Stock Data

In this paper, only announcements published by the cse.lk website was considered. The total data for the past five years is collected.

#### Step 02: Data Preprocessing

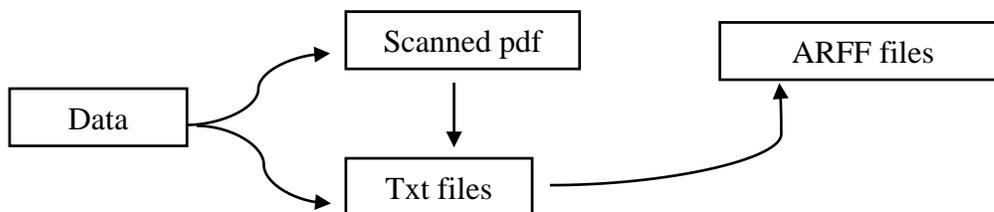


Figure 2: Data Preprocessing

Most of the data collected in cse.lk are scanned files and some are text files. Text files (.txt) can be directly converted to ARFF files. However, the scanned files (.pdf) cannot be converted to ARFF. Scanned files must first be converted to txt files and converted to ARFF files. PDF can be converted to TXT using “Onlineocr.net” website. After converting the txt files, they are divided into three parts: positive, negative, and not change. The researcher then created a training file using WEKA’s Simple CLI.

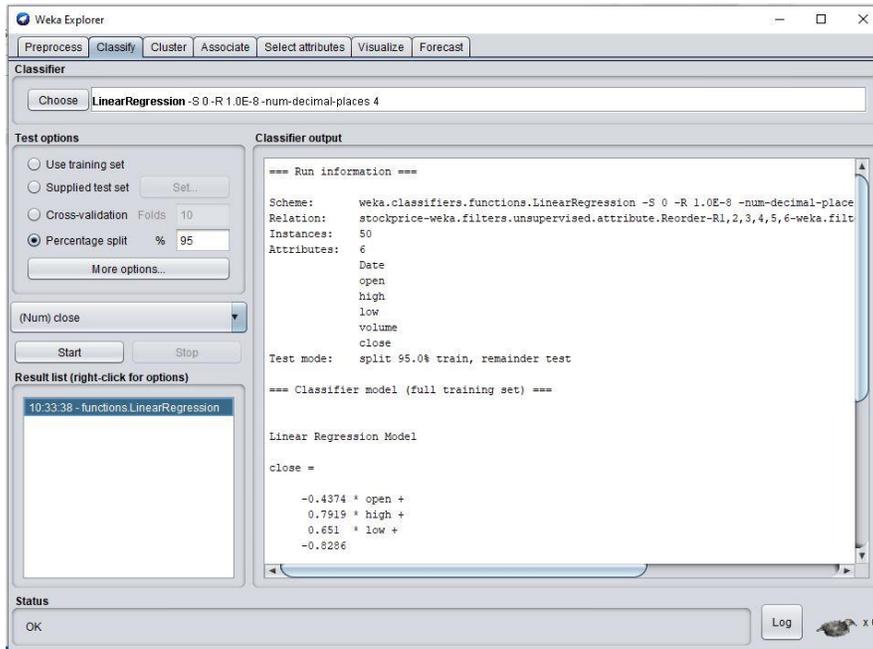
#### Step 03: Learning Algorithm (Learn Rules)

Classification is the process of finding a set of formats that describe and distinguish data classes and concepts so that the label can use the model to predict an unknown class. Classification is a two-step process; first, it creates a classification model using training data.

Each object in the dataset must be preprocessing beforehand, and its class label must be known; second, the model generated in the previous step is tested by assigning class labels to data objects in a test set.

This is about the first step, called training data. It is assumed that each tuple/sample belongs to a predefined class as determined by the class label attribute. In the fundamental analysis, the class attribute was the “Negative”, “Positive” or “Not Change”. The model is represented as a classification rule, decision trees, or mathematical formulas. For the training data, the researcher used a classify tab of WEKA as shown in figure 3.

Figure 3: WEKA Classify Tab



**Step 04: Make Predictions**

There are two steps to the classification process. The first is to build a classification model using training data. Second, the model generated in the previous step is tested using the test dataset. This step discusses how to predict values using the model. It assesses the accuracy of the model. The known label of the test sample is compared with the classification result of the model. Model construction describes a set of predefined classes. The accuracy ratio is the percentage of samples of test kits that are correctly classified by the model. The test set is independent of the training set, otherwise, excessive over-fitting will occur [18].

**Step 5: Evaluate Results**

The accuracy of the model can be estimated by comparing the actual results with the predicted results. Similarly, the Accuracy, Precision, Recall, F-Measure and ROC Area can be evaluated and select the most relevant classification.

**IV. DATA ANALYSIS & PRESENTATION**

*A. Analysis of Data*

For the analysis, the researcher collected data from the CSE's official website. They are textual data, which is an announcement of dividend payments, disclosures, changes to the Directory Board and the Employee Stock Option Schemes. The training dataset looks at the past five years and the test dataset looks at this year's announcements. Table 1 shows how to split the training and test dataset.

Table 1: Text Data Set

Company	Train	Test
---------	-------	------

CTC	43	3
DIAL	71	7
JHK	90	7

The table 2 shows how the training dataset is contained.

Table 2: CTC Training Data Set

Company	Positive	Negative	Not Change	Total
CTC	17	12	14	43
DIAL	33	18	20	71
JKH	42	44	04	90

In the case of CTC, the total number of instances is 43, it contains 17 positive instances, 12 negative instances, and 14 not change instances. The 14 announcements do not affect the stock price. However, 29 announcements affect the company's share price. That is 67% of the total instances. This suggests that announcements can affect the share price. In terms of DIAL, the total number of cases is 71, which includes 33 positives, 18 negatives, and 20 not change cases. The 20 announcements do not affect the stock price. However, 51 announcements affect the company's share price. That is 72% of the total chances. In the case of JKH, there are 90 positive cases, 42 positive, 44 negatives, and four not change. Announcements 4 do not affect the stock price. However, 86 announcements affect the company's share price. That is 95.6% of the total chances. This shows that investors should be very careful about announcements when buying and selling JKH shares.

*B. Analysis of Results*

Before analyzing the data, it must choose a "Filtered Classifier". The researcher can then use any classification and the filter should be "StringToWordVector". The table 3 shows the accuracy of the model using 15-fold, 10-fold, 5-fold cross-validation, and the training set.

Table 3: CTC – J48

Measurement	15 - Folds	10 - Folds	5 - Folds	Use training set
Accuracy (%)	39.5	44.2	34.9	83.7
Precision	0.401	0.465	0.363	0.835
Recall	0.395	0.443	0.349	0.837
ROC	0.525	0.545	0.508	0.939
F - Measure	0.398	0.443	0.351	0.834
AUC	0.525	0.542	0.508	0.940

Use training set" shows higher accuracy. Therefore, the researcher used "use training set" to evaluate the classifiers. The researcher used five classifiers, which are past researchers used. It is as follows,

Table 4 Classifier Evaluation - CTC

Measurement	Naïve Bayes	SMO	J48	Random Forest	Random Tree
Accuracy (%)	90.7	100	83.72	100	100
Precision	0.919	1.000	0.835	1.000	1.000
Recall	0.907	1.000	0.837	1.000	1.000
ROC	0.975	1.000	0.939	1.000	1.000
F - Measure	0.902	1.000	0.834	1.000	1.000
AUC	0.975	1.000	0.940	1.000	1.000

MAE	0.0628	0.2222	0.1597	0.1524	0
RMSE	0.2464	0.2722	0.2826	0.1803	0
RAE (%)	14.27	50.48	36.28	34.62	0
RRSE (%)	52.55	58.03	60.25	38.45	0

When the look at the above tables all companies, and all classifiers shows good results. Its accuracy is more than 80% of all models. Similarly, SMO, Random Forest, and Random Tree classifiers show 100% accuracy and 1.0 Precision, Recall, ROC, AUC, and F-measure. The random tree shows the best results. Its accuracy is 100% and the error is zero. Therefore, this is the best classification for stock price prediction using text analysis. The literature says that the random tree classifier is the best classifier for stock price prediction using text analysis [16]. This research can further confirm that.

The tables below (Table 5, Table 6, and Table 7) show a comparison between the predicted results and the actual results. All predictions conducted using "Random Tree".

Table 5: Results Comparison - CTC

Instance	Predicted Results	Actual Results
1	Positive	Not Change
2	Positive	Positive
3	Positive	Negative

Table 6: Results Comparison - DIAL

Instance	Predicted Results	Actual Results
1	Positive	Not Change
2	Positive	Negative
3	Positive	Positive
4	Positive	Positive
5	Positive	Positive
6	Positive	Not Change
7	Not Change	Not Change

Table 7: Results Comparison - JKH

Instance	Predicted Results	Actual Results
1	Negative	Positive
2	Negative	Positive
3	Negative	Negative
4	Negative	Negative
5	Negative	Negative
6	Positive	Negative
7	Negative	Positive

## V. CONCLUSION & RECOMMENDATIONS

In this analysis, its accuracy is more than 80% of all models. Similarly, SMO, Random Forest, and Random Tree classifiers show 100% accuracy and 1.0 Precision, Recall, ROC, and F-measure. The random tree shows the best results. Its accuracy is 100% and the error is zero. Therefore, this is the best classification for stock price prediction using text analysis. The literature says that the random tree classifier is the best classifier for stock price prediction using text analysis. Financial news affects the closing price of CTC 67%, Dial 72%, and JKH 95.6 percentage.

In this study, the researcher considers three different sectors in CSE. All sectors show the best classifier is the same one. However, other sectors may have different classifiers. Therefore, anyone interested in this field can do other areas to identify classifiers. In addition,

the researcher considers only the announcements posted on the CSE website. Newspaper articles, Twitter, FB posts, and financial reporting can also be considered in the fundamental analysis. Therefore, it might be more accurate. The most important thing for further research is to identify the value of fundamental analysis. That is to say how much the value of the shares will increase or decrease by the fundamental analysis.

## REFERENCES

- [1] M. Hossen, A. N. M. Shareef, M. N. Hasan, and P. Sarani, "Fundamentals Knowledge of Investor and Its Influence on Investment in Capital Market-A Study from Dhaka Stock Exchange," 2014. Accessed: Jan. 18, 2020. [Online]. Available: [www.iiste.org](http://www.iiste.org).
- [2] C. Paper, "Available Online at [www.ijarcs.info](http://www.ijarcs.info) PREDICTING THE STOCK PRICE USING LINEAR REGRESSION," vol. 9, no. 3, pp. 81–85, 2018.
- [3] CSE, "Colombo Stock Exchange," *Introduction of Colombo Stock Exchange*, 2015. <https://www.cse.lk/home.do> (accessed Feb. 04, 2020).
- [4] J. L. Treynor, "Market Value, Time, and Risk," *SSRN Electron. J.*, Aug. 2015, doi: 10.2139/ssrn.2600356.
- [5] "Treynor, J.L. (1961) Market Value, Time and Risk (Unpublished Manuscript). - References - Scientific Research Publishing." [https://www.scirp.org/\(S\(351jmbntvnsjt1aadkpozje\)\)/reference/ReferencesPapers.aspx?ReferenceID=1684640](https://www.scirp.org/(S(351jmbntvnsjt1aadkpozje))/reference/ReferencesPapers.aspx?ReferenceID=1684640) (accessed Feb. 07, 2020).
- [6] E. F. Fama, "American Finance Association Efficient Capital Markets: A Review of Theory and Empirical Work," 1970.
- [7] C. Camerer, G. Loewenstein, and M. Rabin, *Advances in behavioral economics*. Russell Sage Foundation, 2004.
- [8] G. A. W. Griffioen, "Technical Analysis in Financial Markets," *SSRN Electron. J.*, p. 322, 2003, doi: 10.2139/ssrn.566882.
- [9] Y. C. Chan, A. C. W. Chui, and C. C. Y. Kwok, "The impact of salient political and economic news on the trading activity," *Pacific Basin Financ. J.*, vol. 9, no. 3, pp. 195–217, Jun. 2001, doi: 10.1016/S0927-538X(01)00015-4.
- [10] W.-B. Yu, B.-R. Lea, and B. Guruswamy, "A THEORETIC FRAMEWORK INTEGRATING TEXT MINING AND ENERGY DEMAND FORECASTING," 2007. Accessed: Feb. 07, 2020. [Online]. Available: [www.amazon.com](http://www.amazon.com).
- [11] A. Nagar and M. Hahsler, "Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams," *Int. Proc. Comput. ...*, vol. 47, no. Icets, pp. 91–95, 2012, doi: 10.7763/IPCSIT.2012.V47.18.
- [12] Y. Shynkevich, T. M. McGinnity, S. Coleman, and A. Belatreche, "Predicting stock price movements based on different categories of news articles," *Proc. - 2015 IEEE Symp. Ser. Comput. Intell. SSCI 2015*, no. December, pp. 703–710, 2015, doi: 10.1109/SSCI.2015.107.
- [13] A. Samuel, "Machine learning - Wikipedia," 1959. [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning) (accessed Feb. 05, 2020).
- [14] V. H. Shah, "Machine Learning Techniques for Stock Prediction," pp. 1–19, 2007.
- [15] K. H. Sadia, A. Sharma, A. Paul, and S. Sanyal, "Stock Market Prediction Using Machine Learning Algorithms," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 4, pp. 25–31, 2019.
- [16] S. Kalmegh, "Analysis of WEKA Data Mining Algorithm REPTree , Simple Cart and RandomTree for Classification of Indian News," *Int. J. Innov. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 438–446, 2015.
- [17] University of Waikato, "Weka 3 - Data Mining with Open Source Machine Learning Software in Java," 2016. <http://www.cs.waikato.ac.nz/ml/weka/> (accessed Feb. 07, 2020).
- [18] S. Singhal and M. Jena, "A Study on WEKA Tool for Data Preprocessing , Classification and Clustering," *Int. J. Innov. Technol. Explor. Eng.*, vol. 2, no. 6, pp. 250–253, 2013, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.687.799&rep=rep1&type=pdf>.
- [19] K. Joshi, B. H. N., and J. Rao, "S TOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS," 2016.

## AUTHORS

**First Author** – Gayan Wickramaratna, Undergraduate Student, Rajarata University of Sri Lanka. [lhmgw@gmail.com](mailto:lhmgw@gmail.com)

**Second Author** – Hashan Ratnayake, Leturer, Rajarata University of Sri Lanka, [hashanpdn@gmail.com](mailto:hashanpdn@gmail.com)

**Correspondence Author** – Gayan Wickramaratna, [lhmgw@gmail.com](mailto:lhmgw@gmail.com) , [l.h.m.gayan@gmail.com](mailto:l.h.m.gayan@gmail.com), + 9471 2023174