CANCER PREDICTION IN EARLY STAGES USING SUPERVISED LEARNING

MANDALA POOJITHA[1], R YASWITHA[1] ,S CHANDRAHASAN[1], SANGARAJU VAMSI KRISHNA[1],PAIDIMUDDALA RAJESH[1], B RAJA KUMAR[2]

1 *Research Scholar, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India*

2 *Professor, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India*

ABSTRACT

Cancer is a disease characterized by the uncontrolled growth and spread of abnormal cells. Early detection and diagnosis of cancer are essential for successful treatment and management of the disease. Machine learning (ML) is a promising approach that can assist in predicting cancer at an early stage, which can lead to better patient outcomes. Several ML algorithms have been applied to predict cancer in its early stages, including decision trees, support vector machines, neural networks, and random forests. These algorithms can be trained on various types of data, including genomic, proteomic, and imaging data. Genomic data can provide important information about gene expression patterns, DNA mutations, and other molecular features of cancer cells. Proteomic data can provide insights into the protein expression patterns that may be indicative of cancer. Imaging data, such as CT and MRI scans, can also provide valuable information about the presence and extent of cancerous lesions. Overall, ML has shown promise as a tool for predicting cancer in its early stages. As the field of ML continues to evolve and improve, it is likely that these algorithms will become even more accurate and reliable in predicting cancer.

Keyword - *Machine Learning, Random Forest, Logistic Regression, Decision Tree, MLtechniques,evaluation*

1. INTRODUCTION

A Cancer research has undergone a continuous evolution over the last few decades. Scientists used various methods, such as early stage screening, to detect cancer types before they cause symptoms. Furthermore, they have created new strategies for predicting cancer treatment outcomes early on. Large amounts of cancer data have been collected and made available to the medical research community as a result of the introduction of new technologies in the field medicine. However, accurate disease prediction is one of the mostinteresting and difficult tasks for physicians.

As a result, machine learning methods have grow popularity among medical researchers. These techniques can discover and identify patterns and relationships between them, from complex datasets, while they are able to effectively predict future outcomes of a cancer type. Given the significance of personalized medicine and the growing trend on the application of ML techniques, we here present a review of studies that make use of these methods regarding the cancer prediction and prognosis.

In these studies prognostic and predictive features are considered which may be independent of a certain treatment or are integrated in order to guide therapy for cancer patients, respectively. In addition, we discuss the types of ML methods being used, the types of data they integrate, the overall performance of eachproposed scheme while we also discuss their pros and cons. Cancer is a disease that affects millions of people worldwide and is responsible for a significant number of deaths every year.

Early detection of cancer is crucial in improving thprognosis and increasing the chances of successful treatment. Machine learning (ML) algorithms have shown promising results in predicting cancer in its early stages,which can lead to timely intervention and better outcomes for patients. The prediction of cancer using ML involves analyzing large amounts of data and identifying patterns and trends that are indicative of cancer. The data used in ML models can include patient information, such as age, gender, family history, lifestyle habits, and medical history.

2. LITERATURE REVIEW

[1] Authors: Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144: 646–74.

The hallmarks of cancer comprise six biological capabilities acquired during the multistep development of human tumors. The hallmarks constitute an organizing principle for rationalizing the complexities of neoplastic disease. They include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis. Underlying these hallmarks are genome instability , , which generates the genetic diversity that expedites their acquisition, and inflammation, which fosters multiple hallmark functions. Conceptual progress in the last decadehas added two emerging hallmarks of potential generality to this list—reprogramming of energymetabolism and evading immune destruction. In addition to cancer cells, tumors exhibit anotherdimension of complexity: they contain a repertoire of recruited, ostensibly normal cells that contribute to the acquisition of hallmark traits by creating the "tumor microenvironment." Recognition of the widespread applicability of these concepts will increasingly affect the development of new means to treat human cancer.

[2] Authors: Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informat 2006;2:59.

Predictive biomarkers to guide therapy for cancer patients are a cornerstone of precision medicine. Discussed herein are considerations regarding the design and interpretation of such predictive biomarker studies. These considerations are important for both planning and interpreting prospective studies and for using specimens collected from completed randomized clinical trials. Specific issues addressed are differentiation between qualitative and quantitative predictive effects, challenges due to sample size requirements for predictive biomarker assessment, and consideration of additional factors relevant to clinical utility assessment, such as toxicity and cost of new therapies as well as costs and potential morbidities associated with routine use of biomarker-based tests.

[3]Authors: Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. Front Genet 2013;4

In the recent years, circulating microRNAs (miRNAs) have garnered a lot of attentionand interest in the field of disease biomarkers. With characteristics such as high stability, low cost, possibility of repeated sampling and minimal invasiveness, circulating miRNAs are ideal for development into diagnostic tests. There have been many studies reported on the potentialof circulating miRNAs as early detection, prognostic, and predictive biomarkers in cancer. Here, we have reviewed the application of plasma and serum miRNAs as biomarkers for cancer focusing on epithelial carcinomas [prostate, breast, lung, colorectal, and gastric cancer(GC)] and hematological malignancies (leukemia and lymphoma). We have also addressed the common challenges that need to be overcome to achieve a successful bench to bedside transition

[4] Authors: zen k, zhang cy. circulating micrornas: a novel class of bio markers

to diagnose and monitor human cancers. med res rev 2012;32:326–48.

Specific and sensitive non-invasive biomarkers for the detection of human epithelial malignancies are urgently required to reduce the worldwide morbidity and mortality caused by cancer. MicroRNAs (miRNAs) are 19-24 nt noncoding RNAs that are frequently dysregulated in cancer and have shown great promise as tissue-based markers for cancer classification. Once thought to be unstable RNA molecules, miRNAs are now shown to be stably expressed in serum, plasma, urine, saliva, and other body fluids. Moreover, the unique expression patterns of these circulating miRNAs are correlated with certain human diseases, including various types of cancer. Therefore, tumor-derived miRNAs in serum or plasma are emerging as novel blood based fingerprints for the detection of human cancers, especially at an early stage. This review presented newly uncovered cellular and molecular mechanisms of the sources and stability of circulating miRNAs, revealing their great potential as a class of highly specific and sensitive biomarkers for tumor classification and prognostication. Meanwhile, this review also addressed certain critical issues that hinder the wide application of this new approaches.

3. METHODOLOGY

3.1 EXISTING SYSTEM –

In the existing system, implementation of machine learning algorithms is bit complex to build due to the lack of information about the data visualization. Machine learning has beenwidely used in the medical field to improve the diagnosis and treatment of various diseases, including cancer. Early detection of cancer is crucial for successful treatment, and machine learning models have been developed to predict cancer in its early stages. To overcome all this, we use machine learning packages available in the scikit-learn library.

DISADVANTAGES:

- High complexity
- Time consuming.

3.1 PROPOSED SYSTEM-

The Proposed several machine learning models to classify cancer stages . These machine learning systems have the potential to greatly improve the early detection of cancerand ultimately save lives. By analyzing large amounts of patient data, these systems can identify patterns and trends that may not be immediately apparent to human doctors. This allows for earlier detection and more effective treatment of cancer. Therefore, we propose a Logistic Regression and Decision Tree, Extra tree , Adaboost , Lda machine Classifier to predict to the levels

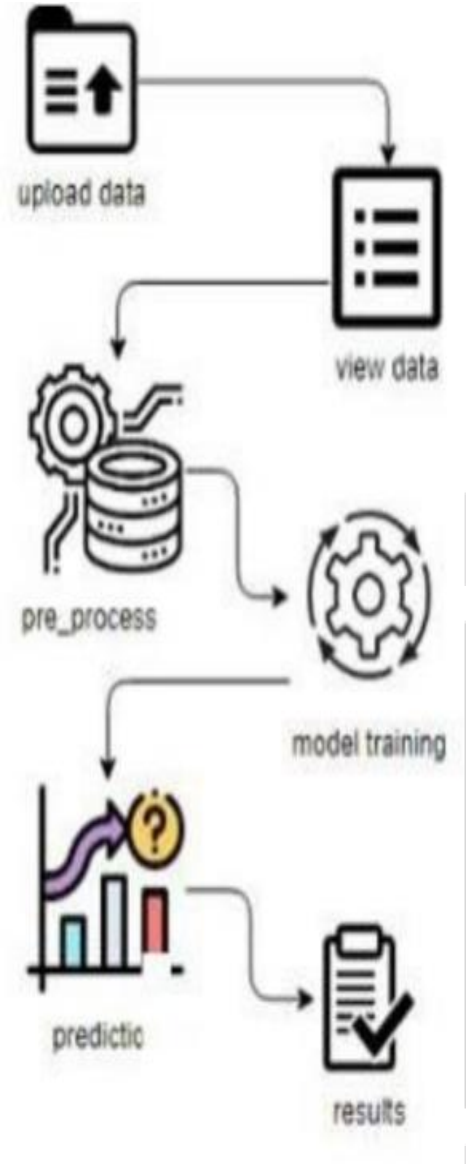


Fig -1 : SYSTEM ARCHITECTUR

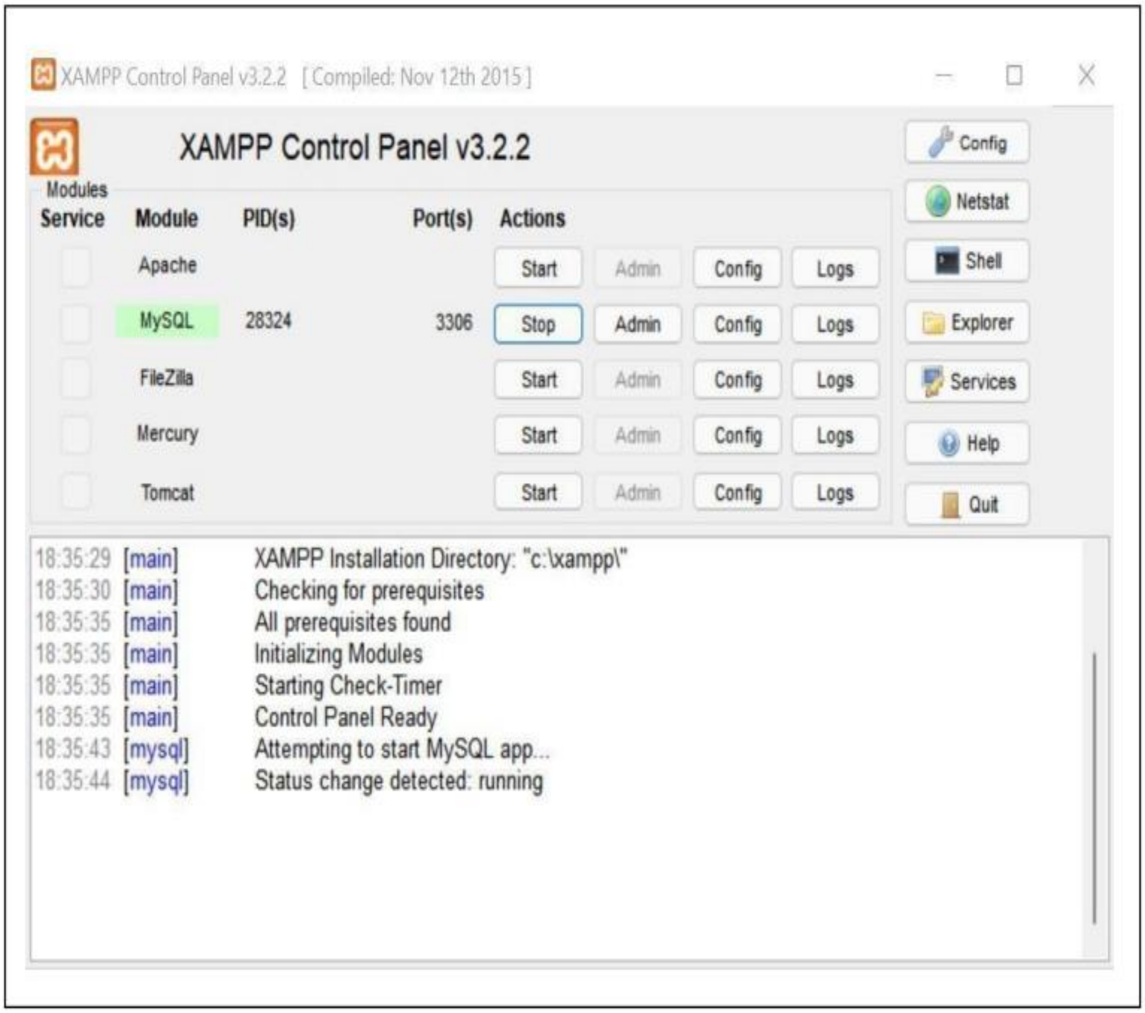
ADVANTAGES:

Highest accuracy.
Reduce time complexity.
Easy to use

3.1 - XAMPP

XAMPP is a free and open-source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database,and interpreters for scripts written in the PHP and Perl programming languages.

XAMPP's ease of deployment means a WAMP or LAMP stack can be installed quickly andsimply on an operating system by a developer, with the advantage that common add-in applications such as Word Press and Joomla! can also be installed with similar ease using Bitnami.


XAMPP Control Panel v3.2.2 [Compiled: Nov 12th 2015]
XAMPP Control Panel v3.2.2
Modules
Config
Netstat
Shell
Explorer
Services
Help
Quit

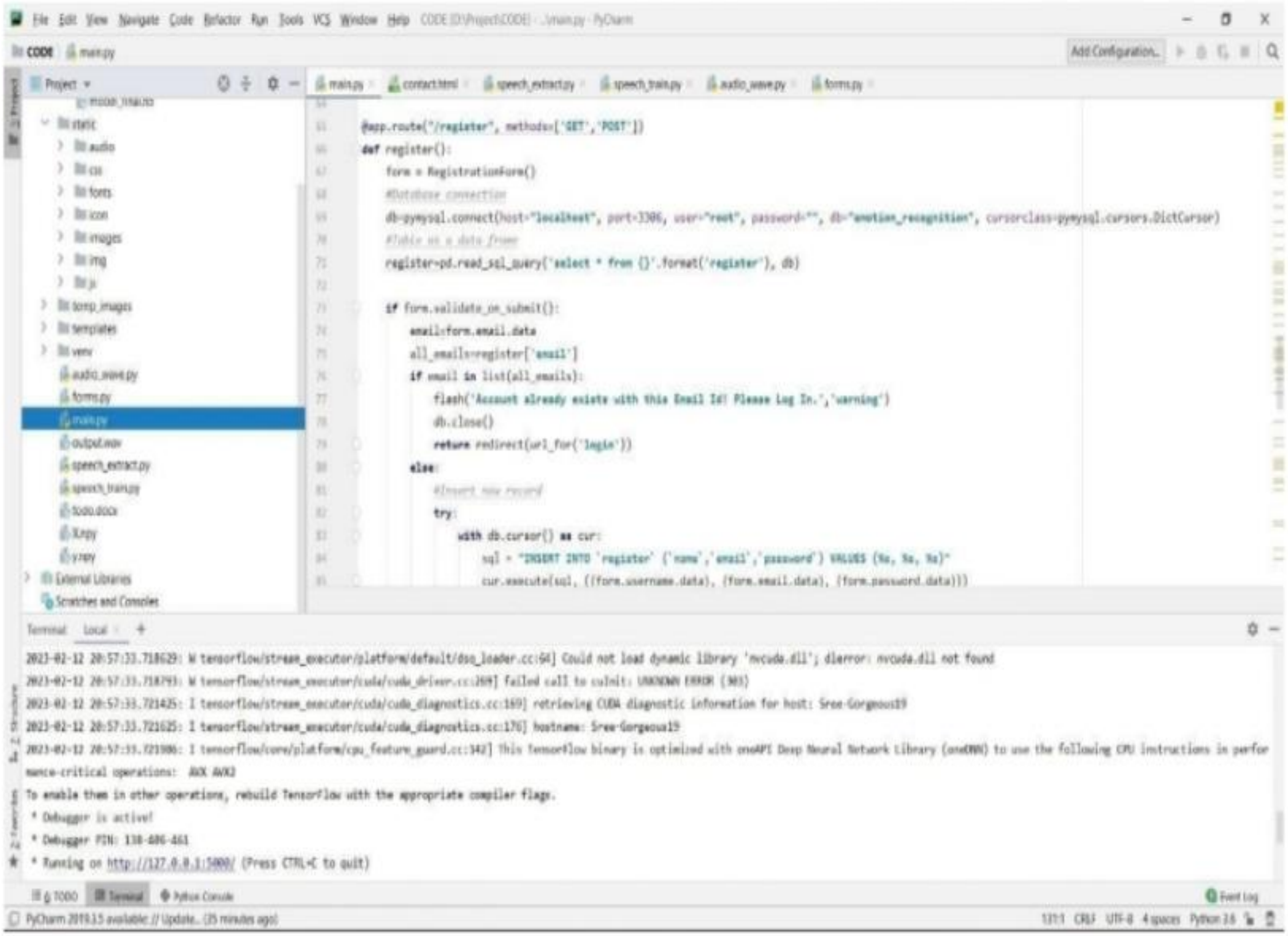
Service	Module	PID(s)	Port(s)	Actions			
□	Apache			Start	Admin	Config	Logs
□	MySQL	28324	3306	Stop	Admin	Config	Logs
□	FileZilla			Start	Admin	Config	Logs
□	Mercury			Start	Admin	Config	Logs
□	Tomcat			Start	Admin	Config	Logs

```
18:35:29  [main]    XAMPP Installation Directory: "c:\xampp\"
18:35:30  [main]    Checking for prerequisites
18:35:35  [main]    All prerequisites found
18:35:35  [main]    Initializing Modules
18:35:35  [main]    Starting Check-Timer
18:35:35  [main]    Control Panel Ready
18:35:43  [mysql]   Attempting to start MySQL app...
18:35:44  [mysql]   Status change detected: running
```

3.1 - PYCHARM

PyCharm is an integrated development environment used in computer programming,specifically for the Python programming language. It is developed by the Czech company JetBrains.

PyCharm is cross-platform, with Windows, macOS and Linux versions. The CommunityEdition is released under the Apache License, and there is also an educational version, as well as a Professional Edition with extra features (released under a subscription- funded proprietary license)



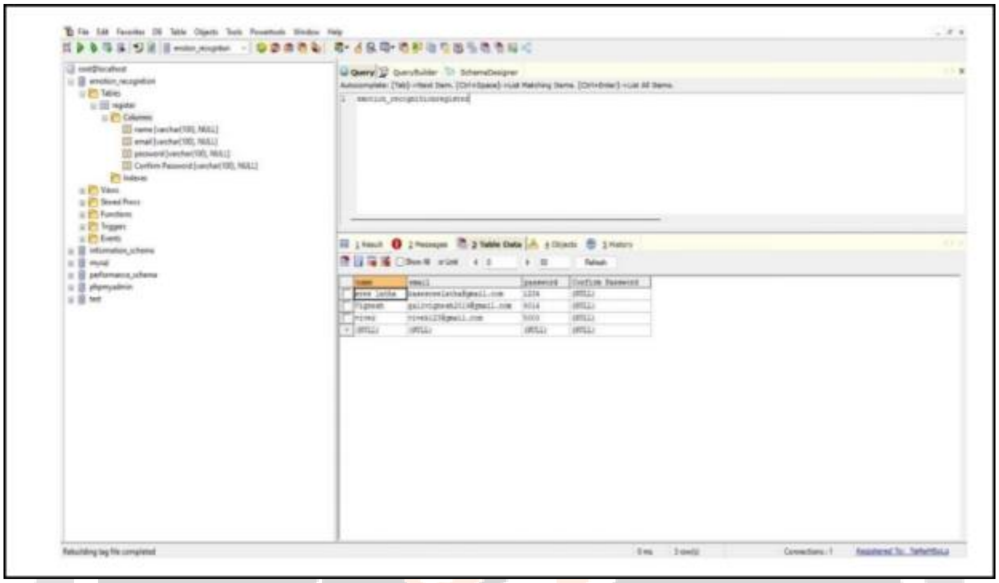
```
@app.route("/register", methods=['GET','POST'])
def register():
    form = RegistrationForm()
    #Database connection
    db=pymysql.connect(host="localhost", port=3306, user="root", password="", db="emotion_recognition", cursorclass=pymysql.cursors.DictCursor)
    #Table as a data frame
    register=pd.read_sql_query('select * from {}'.format('register'), db)

    if form.validate_on_submit():
        email=form.email.data
        all_emails=register['email']
        if email in list(all_emails):
            flash('Account already exists with this Email Id! Please Log In.','warning')
            db.close()
            return redirect(url_for('login'))
        else:
            #Insert new record
            try:
                with db.cursor() as cur:
                    sql = "INSERT INTO `register` (`name`,`email`,`password`) VALUES (%s, %s, %s)"
                    cur.execute(sql, ((form.username.data), (form.email.data), (form.password.data)))
```

```
2023-02-12 20:57:33.718629: W tensorflow/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'nvcuda.dll'; dlerror: nvcuda.dll not found
2023-02-12 20:57:33.718793: W tensorflow/stream_executor/cuda/cuda_driver.cc:269] failed call to cuInit: UNKNOWN ERROR (303)
2023-02-12 20:57:33.721425: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:169] retrieving CUDA diagnostic information for host: Sree-Gorgeous19
2023-02-12 20:57:33.721625: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:176] hostname: Sree-Gorgeous19
2023-02-12 20:57:33.721986: I tensorflow/core/platform/cpu_feature_guard.cc:142] This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in perfor
mance-critical operations:  AVX AVX2
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
 * Debugger is active!
 * Debugger PIN: 138-486-461
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

3.1 : SQLyog

SQLyog SQLyog is a GUI tool for the RDBMS MySQL. It is developed by Webyog, Inc., based in Bangalore, India, and Santa Clara, California. SQLyog is being used by more than 30,000 customers worldwide and has been downloaded more than 2,000,000 times.

SQLyog connects to the MySQL server and stores all the data that has been uploaded to the cloud server.



name	email	password	Confirm Password
[illegible]	[illegible]@gmail.com	1234	(NULL)
[illegible]	[illegible]@gmail.com	[illegible]	(NULL)
[illegible]	[illegible]@gmail.com	[illegible]	(NULL)
(NULL)	(NULL)	(NULL)	(NULL)

Fig- SQLyog

4. MODULES

4.1 USER:

- ***View Home page*** **:** Here user view the home page of the Cancer application.
- **View about page :** In the about page, users can learn more about the cancer platform.
- **View load page :** In the load data page, the user will load the dataset for modeling .
- **View page :** In view page, the user will see the uploaded dataset.
- **Input Model :** The user must provide input values for the certain fields in order to get results.
- **View Result :** User view's the generated results from the model**.**
- **View score :** Here user have ability to view the accuracy score in %

4.1- SYSTEM:

- ***Working on dataset :*** System checks for data whether it is available or not and load the data inexcel files.
- **Pre-processing:**Data need to be pre-processed according the models it helps to increasethe accuracy of the model and better information about the data.
- **Training the data:**After pre-processing the data will split into two parts as train and test databefore training with the given algorithms.
- **Model Building:**To create a model that predicts the personality with better accuracy, thismodule will
- help user**.**
- **Generated Score:**Here user view the score in %.

5. CONCLUSIONS

The implementation of machine learning approaches for covert channel detection demonstrated significant results across various algorithms, including Decision Tree, Random Forest, Logistic Regression, AdaBoost, Extra Trees, and LDA classifiers. Among these, the ExtraTrees algorithm emerged as the most promising method for predicting covert channel attack types.Leveraging diverse ensemble techniques, the study offers insights into the complex patterns of covert communications, providing a robust framework for identifying and mitigating potential security threats. The comparative analysis enhances the understanding of the strengths and weaknesses of different classifiers in the context of cybersecurity.

6. ACKNOWLEDGEMENT

An endeavor of a long period can be successful only with the advice of many well- wishers. I take this opportunity to express my deep gratitude and appreciation to all those who encouraged me for successful completion of the project work.

I am thankful to my Guide **Mr.B.RAJA KUMAR ,M.Tech.,(Ph.D.)** Assistant Professor in Department of **COMPUTER SCIENCE & INFORMATION TECHNOLOGY**, Siddharth Institute of Engineering & Technology, Puttur, for his valuable guidance and suggestions in analyzing and testing throughout the period, till the end of project work.

I wish to thank our Head of the department, **Mr. B Raja Kumar, M.Tech., (Ph.D.)** for his guidance to accomplish the project.

My special thanks to **Dr K Chandra Sekhar Reddy, M.Tech., Ph.D.** Principal, Siddharth Institute of Engineering & Technology, Puttur, for his timely suggestions and help inspite of his busy schedule during the progress of project work.

We extend our heartfelt appreciation to **Dr.P S Raju, Ph.D.** Director of Siddharth Institute of Engineering & Technology, Puttur, for his unwavering support and encouragement that greatly contributed to the successful completion of our project work.

I wish to express my sincere thanks to **Dr K Ashok Raju, Ph.D.** Chairman of SiddharthInstitute of Engineering & Technology, Puttur, for his consistent help and encouragement to complete the project work

.

Finally, I would like to express sincere thanks to Project Co-Ordinator **Mr.C.H.Siva Shankar , M.Tech,.** and **Faculty Members** of CSIT Department and Lab technicians, one and all who have helped me to complete the project work successfully.

As a gesture of respect, I whole heartedly dedicate this work to my **Family Members** and **friends** for their support.

6. REFERENCES

[1] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144: 646-74
[2] Polley M-YC, Freidlin B, Korn EL, Conley BA, Abrams JS, McShane LM. Statistical and practical considerations for clinical evaluation of predictive biomarkers. J Natl Cancer Inst 2013;105:1677– 83.
[3] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informat 2006;2:59.
[4] Fortunato O, Boeri M, Verri C, Conte D, Mensah M, Suatoni P, et al. Assessment of Circulating microRNAs in plasma of lung cancer patients. Molecules 2014;19:3038–54.
[5] Heneghan HM, Miller N, Kerin MJ. MiRNAs as biomarkers and therapeutic targets in cancer.Curr Opin Pharmacol 2010;10:543–50.
[6] Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood

basedcirculating microRNA signatures. Front Genet 2013;4.
[7] Zen K, Zhang CY. Circulating microRNAs: a novel class of biomarkers to diagnose and monitorhuman cancers. Med Res Rev 2012;32:326–48.
[8] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.
[9] Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., ... & Tomaszewski, J. (2014). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Scientific Reports, 7(1), 46450.
[10] Fakoor, R., Ladhak, F., Nazi, A., & Huber, M. (2013). Using deep learning to enhance cancer diagnosis and classification. In Proceedings of the International Conference on Machine Learning (ICML) Workshop on Challenges in Representation Learning (pp. 1-9).
[11] Ting, D. S. W., Cheung, C. Y. L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., ... & Wong, T. Y. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA, 318(22), 2211-2223.
[12] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8-17.