International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

EARLY LUNG CANCER PREDICTION USING CTGAN AND TREE-BASED MACHINE LEARNING TECHNIQUES

Bushra Khanam[1], Lubna Nausheen[2], Faiza Fatima[3]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad,
khanambushra11@gmail.com
[2]Asst Professor, Department of CSE, Shadan Women's College of Engineering and Technology
lubnanausheen07@gmail.com
[3]Asst Professor, Department of CSE, Shadan Women's College of Engineering and Technology
faizafatima010@gmail.com

ABSTRACT
One of the main causes of cancer-related fatalities globally is still lung cancer, and increasing survival rates depends on early identification. In this work, a machine learning-based method for predicting lung cancer utilising clinical and lifestyle data from surveys is presented. The Synthetic Minority Over-sampling Technique (SMOTE) was used to address the dataset's class imbalance and guarantee equitable representation of both classes. Logistic regression was used as the meta-learner in a stacked ensemble model that combined CatBoost, XGBoost, LightGBM, AdaBoost, and Random Forest. The suggested methodology outperformed individual classifiers and came close to state-of-the-art performance documented in the literature, with an accuracy of 96.9% and a ROC AUC of 0.99. A Flask-based web application that offers an intuitive interface for prediction, visualisation, and outcome interpretation was also developed. Modules for user registration, input-based cancer risk prediction, and graphical display of test data results are all included in the system. The findings show that ensemble learning provides a reliable and user-friendly method for lung cancer prediction when combined with efficient preprocessing and web deployment.

1. INTRODUCTION
The concept of artificial intelligence (AI) has revolutionised a number of fields, including engineering, manufacturing, and medicine. In particular, the development of automated solutions in medical science has been considerably aided by the application of machine learning techniques. Medical professionals can create specialised and focused treatment plans by using machine learning models to reliably and accurately analyse massive streams of medical data. Lung cancer is one of the top causes of cancer-related fatalities globally. Early diagnosis and treatment greatly increase patient survival and lower death rates. With 40.9 million patients, lung cancer was the leading cause of disability-adjusted life years (DALYs) in 2017.

Lung cancer is thought to have caused up to 1.8 million deaths in 2020 [1], [2], and statistics indicate that up to 135,720 patients would die from the disease in the United States alone. Nevertheless, lung cancer does not exhibit symptoms at first and then becomes difficult to detect at an early stage. The current work employs machine learning (ML) and deep learning (DL) techniques to analyse and forecast the development and course of lung cancer [3]. Small Cell Lung Cancer (SCLC) accounts for 10–16% of lung cancer cases. SCLC develops at a very quick pace, which eventually leads to the growth of tumours that frequently spread widely and are closely linked to smoking cigarettes. Through the study of enormous volumes of patient records, machine learning (ML) algorithms may accurately predict whether a patient will acquire lung cancer at an early stage by identifying disease patterns and characteristics associated with the condition. Lung cancer can now be detected early thanks to recent advancements in bioinformatics (images and factor-based). possible. The idea is to predict a patient's risk by extracting relevant features from these datasets and applying a robust ML algorithm. ML algorithms can significantly improve the accuracy of malignancy predictions. In [4], authors have applied several metrics like Confusion Matrix, Precision, F1 Score, Recall, and Accuracy, along with multiple algorithmic approaches such as Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbor (KNN). All these techniques are used to recognize gene expression signatures from samples and develop predictive cancer models. The Authors designed and evaluated an ADB-based prognostic model to provide medical professionals with an ECOG PS-based decision-making process for lung cancer patients. Many scientific articles have addressed the application of ML in this particular disease. We propose a methodology for designing effective ML classification models for lung cancer prediction utilizing everyday routines and symptoms/signs as input characteristics for the models. AI and ML play significant role in modern healthcare. The efficacy of ML models is heavily reliant on how effectively the model can classify true and false positives and negatives based on the different datasets used for classification [5]. The study focuses on precision,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

accuracy, and F1 scores, considering a significant body of literature. The results of the current study discussed next, represent significant improvements compared with earlier studies [6], [7]. In particular, it will assess the efficiency of different classifiers like XGBoost, Extra Trees Classifier (ETC), SVM, KNN, Stochastic-Gradient-Descent Classifier (SGDC), LR, DT, RF, and Gradient-Boosting Classifier (GBC) for early-detection of cases with lung cancer by analyzing the pertinent medical data [8]. The purpose of the review is to go over studies that employ clinical risk factor data and various methods to predict lung cancer. It contains an overview of feature extraction and selection methods, ML and DL algorithms with various validation measures, and survival prediction procedures. Additionally, a homemade e-nose device for lung cancer detection has been proposed by certain researchers [9], 34322 [10]. There is still a strong incentive to improve the categorisation techniques despite all the effort that has been put into this field. Additionally, neural networks have been employed to improve classification performance; feedforward and backpropagation techniques are particularly popular in these endeavours [11], [12].

2. EXISTING SYSTEM OF PROJECT

To enhance lung cancer detection and classification, the current approach in this project combines tree-based learning models, particularly the Random Forest classifier, with Conditional Tabular Generative Adversarial Networks (CTGAN). The problem of class imbalance that is frequently present in medical datasets is addressed by using CTGAN to create artificial tabular data. This improves the training data's representativeness and variety.

Then, predictive models that can differentiate between various patient situations are built using the Random Forest algorithm, which is renowned for its robustness and ensemble-based approach. In order to develop a more balanced, accurate, and efficient diagnostic system for early lung cancer diagnosis, this integrated strategy combines the creation of synthetic data with dependable decision-tree-based classification.

ALGORITHM:

Conditional Tabular Generative Adversarial Network (CTGAN) is a deep learning model designed to generate realistic synthetic tabular data by learning the underlying distribution of the original dataset. It helps to address issues like class imbalance and data scarcity, especially in sensitive domains such as healthcare, by producing high-quality, representative samples that augment the training data. Random Forest (RF), on the other hand, is an ensemble machine learning algorithm that constructs multiple decision trees during training and combines their outputs to improve classification accuracy and reduce overfitting. By leveraging the diversity of trees, RF offers robust and reliable predictions even in complex datasets. Together, CTGAN and RF form a powerful combination where CTGAN enhances the dataset quality and balance, while RF effectively performs classification tasks, leading to improved model performance in applications like early lung cancer detection.

Drawbacks: -

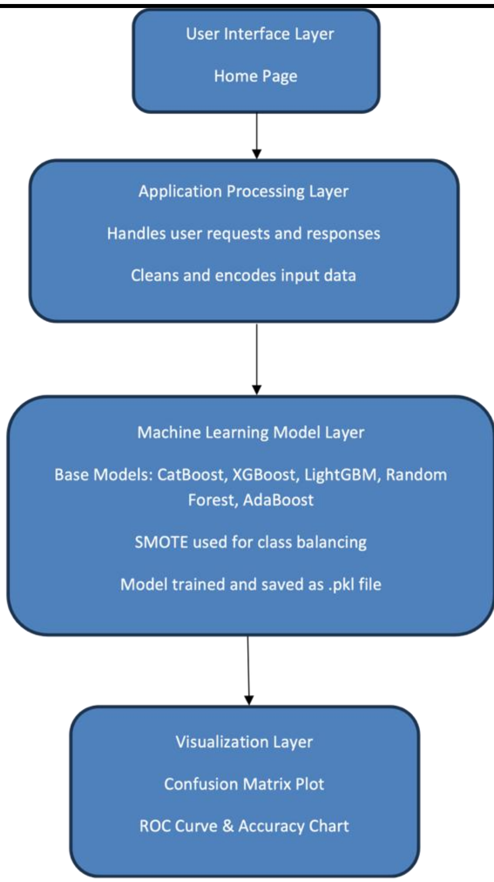
- Computational Complexity
- Synthetic Data Quality Dependency
- Interpretability Challenges

3. PROPOSED SYSTEM

In order to improve lung cancer prediction accuracy, robustness, and reliability, the suggested method presents a Stacked Ensemble Learning Model that combines several sophisticated classifiers. The suggested method creates a multi-layered ensemble by combining the predictive strengths of multiple boosting algorithms, including Random Forest, CatBoost, XGBoost, LightGBM, and AdaBoost, rather than relying on a single model. The last layer uses a Logistic Regression meta-learner to combine base model predictions and provide the final choice. The Synthetic Minority Over-sampling Technique (SMOTE) is used to balance the dataset and guarantee equitable representation of both cancer and non-cancer cases in order to further improve performance. Flask is used in the deployment of this system, allowing for an accessible and interactive web interface for real-time prediction and visualisation.

SYSTEM ARCHITECTURE:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



ALGORITHM:

Stacked Ensemble Learning Model (CatBoost + XGBoost + LightGBM + AdaBoost + Random Forest with Logistic Regression Meta-Learner)

The stacked ensemble model is a two-layer machine learning framework. In the first layer, multiple base classifiers are trained independently on the same dataset to generate diverse predictions. In the second layer, a meta-learner (Logistic Regression) takes these predictions as input and learns how to optimally combine them to produce a more accurate final result. This method leverages the strengths of multiple algorithms, minimizes their individual weaknesses, and provides superior predictive performance compared to single-model systems.

Advantages: -

- Robustness
- Better Generalization
- High Accuracy

4. DATASET PREPARATION:

This paper's dataset was retrieved from the public repository Kaggle [27]. It contains 309 instances and 16 attributes: 15 predictive attributes, and 1 class attribute. Class attributes are lung cancer, and predictive attributes include Gender, Age, Anxiety, Smoking, Peer Pressure, Yellow Fingers, Chronic Disease, Allergy, Fatigue, Alcohol, Wheezing, Coughing, Swallowing Difficulty, Lung Cancer, Shortness of Breath, Chest Pain. This dataset is highly imbalanced, as 270 records are of the 'cancer' class while 39 belong to the 'normal' class. Class imbalance introduces bias in the results, significantly affecting the prediction process. The table 2 describes each aspect of the dataset.

TECHNIQUES USED IN PROJECT

The proposed lung cancer prediction system utilizes several advanced machine learning and data processing techniques to improve prediction accuracy and reliability. Initially, data preprocessing techniques are applied to clean the dataset, handle missing values, and convert categorical attributes into a machine-readable format. To address the issue of class imbalance commonly found in medical datasets, the Synthetic Minority Over-sampling Technique (SMOTE) is employed, which generates synthetic samples for the minority class and ensures balanced training data. The core of the system is based on a stacked ensemble learning approach that combines multiple powerful machine learning algorithms, including CatBoost, XGBoost, LightGBM, AdaBoost, and Random Forest. Each of these classifiers learns different patterns from the dataset and produces individual predictions. CatBoost efficiently handles categorical features, XGBoost provides optimized gradient boosting with regularization, LightGBM offers fast and memory-efficient training, AdaBoost improves classification by focusing on misclassified instances, and Random Forest enhances robustness through multiple decision trees. The predictions generated by these base models are then combined using Logistic Regression as a meta-learner, which learns the optimal way to integrate the outputs of all classifiers and generate the final prediction. This stacking strategy improves overall accuracy, reduces overfitting, and enhances generalization capability. The model's performance is evaluated using metrics such as Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC Score to ensure reliable classification results. Finally, the trained model is deployed using the Flask web framework, enabling users to input clinical information, obtain real-time lung cancer predictions, and visualize results through an interactive web interface. This combination of data balancing, ensemble learning, performance evaluation, and web deployment techniques results in an efficient and accurate system for early lung cancer detection.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

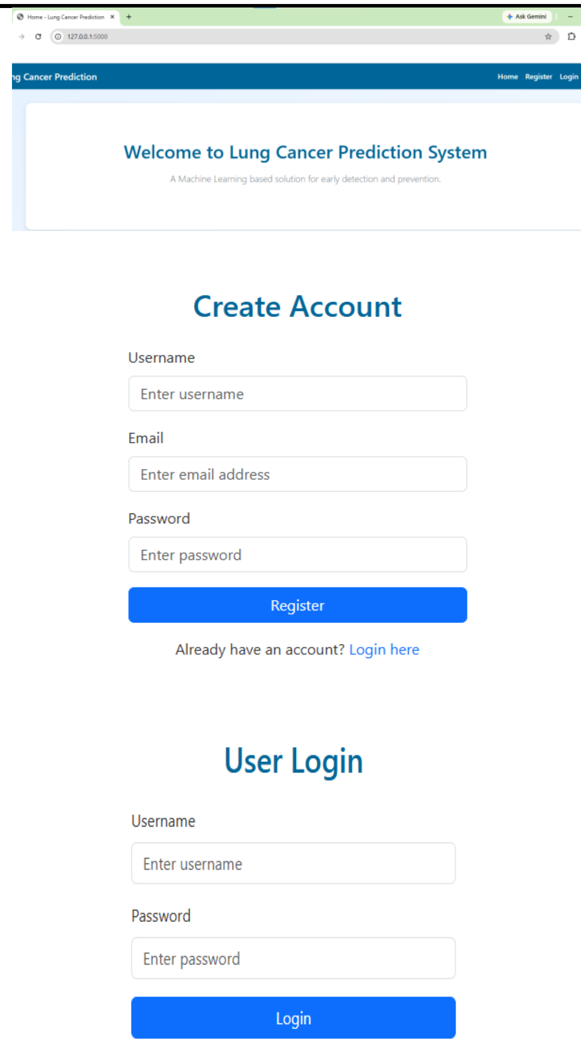
5. MODEL TRAINING

The gathering and preprocessing of the lung cancer dataset is the first step in the model training process. To eliminate inconsistencies and get the dataset ready for machine learning analysis, data cleaning techniques are used. Given the prevalence of class imbalance in medical records, the Synthetic Minority Over- To provide a balanced class distribution and enhance the model's capacity to learn from all categories, the sampling Technique (SMOTE) is used to create synthetic samples for the minority class. After that, the preprocessed dataset is split into training and testing sets so that the model can assess its performance on untested data while learning patterns from the training data.

A stacked ensemble learning strategy is used to attain excellent prediction accuracy. The balanced training dataset is used to individually train five basic classifiers in the first layer: CatBoost, XGBoost, LightGBM, AdaBoost, and Random Forest. Different traits and connections between the lifestyle and clinical variables linked to lung cancer are learned by each model. AdaBoost concentrates on hard-to-classify samples, Random Forest increases resilience through multiple decision trees, XGBoost and LightGBM offer optimised gradient boosting capabilities, and CatBoost effectively manages categorical data. The second layer of the ensemble architecture uses the predictions produced by these foundation models as input characteristics.

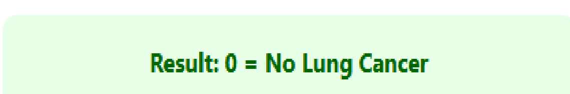
In the second layer, Logistic Regression is employed as a meta-learner to combine the outputs of the base classifiers. The meta-learner analyzes the prediction patterns of each model and learns the optimal combination strategy to produce the final classification result. This stacking technique reduces the limitations of individual models and enhances overall predictive performanceHyperparameters are adjusted during training to enhance model generalisation and avoid overfitting. Several performance indicators, such as Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC-AUC Score, are then used to assess the trained ensemble model. With an accuracy of roughly 96.9% and a ROC-AUC score of 0.99, experimental data show that the stacked ensemble model performs better and is useful for early lung cancer prediction. For real-time prediction and visualisation, the finished trained model is stored and incorporated into a Flask-based web application..

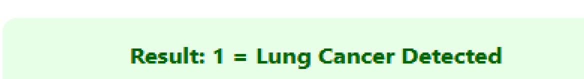
6. RESULT:



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper


Lung Cancer Prediction Form
Gender (0 = Female, 1 = Male)
Female
Age
Smoking (1 = Yes, 2 = No)
Yes
Yellow Fingers (1 = Yes, 2 = No)
Yes
Anxiety (1 = Yes, 2 = No)
Yes
Peer Pressure (1 = Yes, 2 = No)
Yes
Chronic Disease (1 = Yes, 2 = No)
Yes
Fatigue (1 = Yes, 2 = No)
Yes
Allergy (1 = Yes, 2 = No)
Yes
Wheezing (1 = Yes, 2 = No)
Yes
Alcohol Consumption (1 = Yes, 2 = No)
Yes
Coughing (1 = Yes, 2 = No)
Yes
Shortness of Breath (1 = Yes, 2 = No)
Yes
Swallowing Difficulty (1 = Yes, 2 = No)
Yes
Chest Pain (1 = Yes, 2 = No)
Yes





As for the experiment performed in the environment of lung cancer detection, this set was also split into a training set containing 80% of data and a test set containing 20% of data. It did have a standard distribution strategy in place to stop it from overfitting the data. In testing the performance of the model, all the metrics some of these include accuracy, precision, Recall, and F1 Score. The experiments were executed in a Python environment which contains many libraries on a powerful server, Dell PowerEdge T430 with a graphic processing unit. This model of GPU has 2GB RAM and is driven by two Intel's Xeon processors each having 8 processing cores and functioning at a rate of 2.4 GHz supported by 32 GB of DDR4 memory. These resources ensured reliable computational performance for the analytical and validation procedures of the formulated models for diagnosing lung cancer.
The proposed stacked ensemble learning model, which combines CatBoost, XGBoost, LightGBM, AdaBoost, and Random Forest with Logistic Regression as the meta-learner, achieved outstanding performance in lung cancer prediction. After applying SMOTE to balance the dataset and training the ensemble framework, the model obtained an accuracy of 96.9% and a ROC-AUC score of 0.99. The high accuracy demonstrates the model's ability to correctly classify both lung cancer and non-lung cancer cases, while the ROC-AUC score indicates excellent discrimination capability between the two classes. The integration of multiple classifiers enabled the model to capture diverse patterns within the data, reducing prediction errors and improving overall reliability. The proposed model exhibited strong generalization performance on unseen test data, confirming its effectiveness for early lung cancer detection. These results highlight the superiority of the stacked ensemble approach over traditional single-model techniques and demonstrate its potential as a robust decision-support tool in healthcare applications.

7. DISCUSSION AND CONCLUSION

The experimental results demonstrate the effectiveness of the proposed stacked ensemble learning model for early lung cancer prediction. The framework integrates CatBoost, XGBoost, LightGBM, AdaBoost, and Random Forest as base classifiers, while Logistic Regression acts as the meta-learner to generate the final prediction. To address the issue of class imbalance commonly found in medical datasets, the Synthetic Minority Over-sampling Technique (SMOTE) was applied during the preprocessing stage, ensuring balanced representation of cancer and non-cancer cases.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The ensemble architecture effectively combines the strengths of multiple machine learning algorithms, enabling the model to learn diverse patterns and relationships within the clinical and lifestyle data. As a result, the proposed system achieved an accuracy of 96.9% and a ROC-AUC score of 0.99, demonstrating excellent predictive capability and strong discrimination between classes. The confusion matrix analysis revealed a low number of misclassifications, indicating reliable and consistent performance. Furthermore, the ensemble model outperformed traditional single-model approaches by reducing prediction errors, improving robustness, and enhancing generalization on unseen data. The deployment of the trained model through a Flask-based web application further validates its practical applicability by enabling real-time prediction, visualization, and user interaction.

In conclusion, this study presents an efficient and reliable lung cancer prediction system based on stacked ensemble learning. By combining multiple advanced classifiers and utilizing SMOTE for data balancing, the proposed model successfully improves prediction accuracy and overall classification performance. The obtained results confirm that ensemble learning can significantly enhance the early detection of lung cancer, providing a valuable decision-support tool for healthcare applications. The integration of the predictive model into a web-based platform further increases its accessibility and usability in real-world environments. Therefore, the proposed system offers a promising approach for assisting medical professionals in the timely identification of high-risk patients and supporting informed clinical decision-making. Future research may focus on incorporating larger and more diverse clinical datasets, integrating deep learning techniques, and implementing explainable artificial intelligence methods to further improve model performance, interpretability, and practical deployment in healthcare settings.

8. FUTURE ENHANCEMENT

Even though the suggested stacked ensemble learning model shown strong performance and high accuracy in lung cancer prediction, a number of improvements could be added in subsequent studies to increase its efficacy and usefulness. Using larger and more varied clinical datasets gathered from other healthcare facilities is one possible enhancement that would improve the model's capacity for generalisation and dependability across different populations. Advanced deep learning architectures such as Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Transformer-based models can also be integrated with the existing ensemble framework to capture more complex patterns within medical data. Furthermore, Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME can be incorporated to provide transparent and interpretable predictions, helping healthcare professionals better understand the factors influencing the model's decisions.

In order to create a multimodal lung cancer prediction system, future research may also examine the integration of clinical and lifestyle data with medical imaging data, such as CT scans, X-rays, and histopathological pictures. To increase accessibility and facilitate remote patient monitoring, real-time cloud deployment and mobile healthcare apps can be used. To further improve prediction accuracy and computing efficiency, sophisticated feature selection approaches and automated hyperparameter optimisation procedures can be applied. It is possible to sustain performance over time and adjust to evolving clinical trends by continuously retraining the model with fresh patient data. The suggested system could become a comprehensive, sophisticated, and clinically useful decision-support tool for early lung cancer detection and diagnosis with these improvements.

REFERENCES

[1] R. Canetti, U. Feige, O. Goldreich, and M. Naor, ''Adaptively secure multi-party computation,'' in *Proc. 28th Annu. ACM Symp. Theory Com- put.*, Philadelphia, PA, USA, 1996, pp. 639–648. [Online]. Available: http://portal.acm.org/citation.cfm?doid=237814.238015

[2] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, ''Deniable encryption,'' in *Proc. 17th Annu. Int. Cryptol. Conf.* (Lecture Notes in Computer Sci- ence), vol. 1294. Santa Barbara, CA, USA: Springer, 1997, pp. 90–104.

[3] A. Juels and T. Ristenpart, ''Honey encryption: Security beyond the brute-force bound,'' in *Advances in Cryptology—EUROCRYPT* (Lec- ture Notes in Computer Science), vol. 8441, D. Hutchison, T. Kanade,

[4] J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, P. Q. Nguyen, and E. Oswald, Eds. Berlin, Germany: Springer, 2014, pp. 293–310, doi: 10.1007/978-3-642-55220-5_17.

[5] M. Durmuth and D. M. Freeman, ''Deniable encryption with negli- gible detection probability: An interactive construction,'' in *Advances in Cryptology—EUROCRYPT* (Lecture Notes in Computer Science), vol. 6632, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern,

[6] J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, and K. G. Paterson, Eds. Berlin, Germany: Springer, 2011, pp. 610–626, doi: 10.1007/978-3-642-20465-4_33.

[7] A. O'Neill, C. Peikert, and B. Waters, ''Bi-deniable public-key encryption,'' in *Advances in Cryptology—CRYPTO* (Lecture Notes in Computer Science), vol. 6841, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg,

[8] F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. P. Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, and P. Rogaway, Eds. Berlin, Germany: Springer, 2011, pp. 525–542, doi: 10.1007/978-3-642-22792-9_30.

[9] M. Klonowski, P. Kubiak, and M. Kutylowski, ''Practical deniable encryp- tion,'' in *SOFSEM 2008: Theory and Practice of Computer Science* (Lecture Notes in Computer Science), V. Geffert, J. Karhumäki, A. Bertoni,B. Preneel, P. Návrat, and M. Bieliková, Eds. Berlin, Germany: Springer, 2008, pp. 599–609.

[10]P. Gasti, G. Ateniese, and M. Blanton, ''Deniable cloud storage: sharing files via public-key deniability,'' in *Proc. 9th Annu. ACM Workshop Pri- vacy Electron. Soc.*, Chicago, IL, USA, 2010, p. 31. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1866919.18669 25

[11] M. H. Ibrahim, ''A method for obtaining deniable public-key encryption,'' *Int. J. Netw. Secur.*, vol. 8, no. 1, pp. 1–9, 2009.

[12] P. Chi and C. Lei, ''Audit-free cloud storage via deniable attribute- based encryption,'' *IEEE Trans. Cloud Comput.*, vol. 6, no. 2, pp. 414–427, Apr. 2018. [Online]. Available: https://ieeexplore.ieee. org/document/7090980/

[13] R. Canetti, S. Park, and O. Poburinnaya, ''Fully deniable interactive encryption,'' in *Advances in Cryptology—CRYPTO* (Lecture Notes in Computer Science), vol. 12170, D. Micciancio and T. Ristenpart, Eds. Cham, Switzerland: Springer, 2020, pp. 807–835, doi: 10.1007/978-3-030- 56784-2_27.

[14] C. Dwork, M. Naor, and A. Sahai, ''Concurrent zero-knowledge,'' *J. ACM*, vol. 51, no. 6, p. 851–898, Nov. 2004, doi: 10.1145/1039488.1039489.

[15] M. Naor, ''Deniable ring authentication,'' in *Advances in Cryptology— CRYPTO* (Lecture Notes in Computer Science), vol. 2442, G. Goos,

[16] J. Hartmanis, J. van Leeuwen, and M. Yung, Eds. Berlin, Germany: Springer, 2002, pp. 481–498, doi: 10.1007/3-540-45708-9_31.

[17] R. W. Zhu, D. S. Wong, and C. H. Lee, ''Cryptanalysis of a suite of deniable authentication protocols,'' *IEEE Commun. Lett.*, vol. 10, no. 6, pp. 504–506, Jun. 2006.

[18] A. Fiat and M. Naor, ''Broadcast encryption,'' in *Proc. Annu. Int. Cryptol. Conf.*, Jan. 1993, pp. 480–491.

[19] J. Li, Y. Wang, Y. Zhang, and J. Han, ''Full verifiability for out- sourced decryption in attribute based encryption,'' *IEEE Trans. Services Comput.*, vol. 13, no. 3, pp. 478–487, May 2020. [Online]. Available: https://ieeexplore.ieee.org/document/7936626/

[20] J. Li, Q. Yu, and Y. Zhang, ''Hierarchical attribute based encryption with continuous leakage-resilience,'' *Inf. Sci.*, vol. 484, pp. 113–134, May 2019. [Online]. Available: https://linkinghub.elsevier. com/retrieve/pii/S0020025519300684

[21] J. Li, W. Yao, Y. Zhang, H. Qian, and J. Han, ''Flexible and fine-grained attribute-based data storage in cloud computing,'' *IEEE Trans. Services Comput.*, vol. 10, no. 5, pp. 785–796, Sep. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7390098/

[22] S. Reddy, P. S. Reddy, and P. Sravanthi, ''Audit free cloud stor- age via deniable attribute base encryption for protecting user pri- vacy,'' *Int. J. Sci. Eng. Technol. Res.*, vol. 5, no. 17, pp. 3449–3451, 2016.

[23] R. Bassous, R. Bassous, H. Fu, and Y. Zhu, ''Ambiguous multi-symmetric cryptography,'' in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 7394–7399.

[24] A. Chakraborti, C. Chen, and R. Sion, ''POSTER: DataLair: A storage block device with plausible deniability,'' in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1757–1759. [Online]. Available: https://dl.acm.org/doi/10.1145/2976749. 2989061

[25] C. Chen, A. Chakraborti, and R. Sion, ''PD-DM: An efficient locality-preserving block device mapper with plausible deniability,'' *Proc. Privacy Enhancing Technol.*, vol. 2019, no. 1, pp. 153–171, Jan. 2019. [Online]. Available: https://petsymposium.org/popets/2019/ popets-2019-0009.php

[26] C. Chen, X. Liang, B. Carbunar, and R. Sion, ''SoK: Plausibly deni- able storage,'' *Proc. Privacy Enhancing Technol.*, vol. 2022, no. 2, pp. 132–151, Apr. 2022. [Online]. Available: https://petsymposium. org/popets/2022/popets-2022-0039.php

[27] E.-O. Blass, T. Mayberry, G. Noubir, and K. Onarlioglu, ''Toward robust hidden volumes using write-only oblivious RAM,'' in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 203–214, doi: 10.1145/2660267.2660313.

[28] C. Hargreaves and H. Chivers, ''Detecting hidden encrypted vol- umes,'' in *Communications and Multimedia Security*, B. De Decker and I. Schaumuller-Bichl, Eds. Berlin, Germany: Springer, 2010, pp. 233–244.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper