IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 July 2023, accepted 15 July 2023, date of publication 20 July 2023, date of current version 26 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297498

RESEARCH ARTICLE

Classification of Potentially Hazardous Asteroids Using Supervised Quantum Machine Learning

RUSHIR BHAVSAR[1], NILESH KUMAR JADAV[1], (Student Member, IEEE), UMESH BODKHE[1], RAJESH GUPTA[1], (Member, IEEE), SUDEEP TANWAR[1], (Senior Member, IEEE), GULSHAN SHARMA[2], PITSHOU N. BOKORO[2], (Member, IEEE), AND RAVI SHARMA[3]

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat 382481, India
[2]Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa
[3]Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun 248001, India

Corresponding authors: Rajesh Gupta (rajesh.gupta@nirmauni.ac.in) and Sudeep Tanwar (sudeep.tanwar@nirmauni.ac.in)

ABSTRACT Quantum computing (QC) and quantum machine learning (QML) are emerging technologies with the potential to revolutionize the way we approach complex problems in mathematics, physics, and other fields. The increasing availability of data and computing power has led to a rise in using Artificial Intelligence (AI) to solve real-time problems. In space science, employing AI-based approaches to address various challenges, including the potential risks posed by asteroids, is becoming increasingly necessary. Potentially Hazardous Asteroids (PHAs) can cause significant harm to humans and biodiversity through wind blasts, overpressure shock, thermal radiation, cratering, seismic shaking, ejecta deposition, and even tsunamis. Machine Learning (ML) algorithms have been employed to detect hazardous asteroids based on their parameters. Still, there are limitations to the current techniques, and the results have reached a saturation point. To address this issue, we propose a Quantum Machine Learning (QML)-based approach for asteroid hazard prediction, employing Variational Quantum Circuits (VQC) and PegasosQSVC algorithms. The proposed work aims to leverage the quantum properties of the data to improve the accuracy and precision of asteroid classification. Our study focuses on the impact of PHAs, and the proposed supervised QML-based method aims to detect whether an asteroid with specific parameters is hazardous or not. We compared several classification algorithms and found that the proposed QML-based approach employing VQC and PegasosQSVC outperformed the other methods, with an accuracy of 98.11% and an average F1-score of 92.69%.

INDEX TERMS Quantum computing, quantum machine learning, qubits, quantum gates, entanglement, astronomy, asteroid hazard prediction, astrometry, quantum algorithm optimization.

I. INTRODUCTION

The universe has always been a subject of fascination for humans. The vastness and complexity of the cosmos have led to numerous questions and theories that have been pondered for centuries. Astronomy, the scientific study of celestial objects and phenomena, has been an essential tool for unraveling the mysteries of the universe. Reference [1] not only has it allowed us to understand the workings of our solar system, the formation of galaxies, and the birth and death of stars, but the idea of exploring the unknown and discovering new worlds has been a driving force behind many scientific and technological advancements. Over the years, we have sent numerous probes and rovers to explore planets and moons in our solar system. These missions have given us information about the geology, atmosphere, and potential for life on these celestial bodies.

The nearest possible sources for origins study are disparate floating rogue objects, such as asteroids and meteorites, in the Kuiper [2] and Inner Main Belt [3]; some of the most important celestial objects providing valuable insights into the solar system's formation. These objects are essentially the building blocks of planets and can provide us with a glimpse into the early stages of the solar system's formation. They are

The associate editor coordinating the review of this manuscript and approving it for publication was Siddhartha Bhattacharyya.

remnants of the primordial material in the early solar system and have remained unchanged since then. Despite their scientific significance, asteroids and meteorites also threaten our planet significantly. The discovery rate of such objects has increased in the last decade, with the current detection rate of Fast Moving Objects (FMOs), according to the Zwicky Transient Facility (ZTF), having been 100 per year [4]. While the assertion that celestial object impacts could have cataclysmic consequences on Earth may seem alarmist, the potential for significant damage has been substantiated through historical precedent. One such event was the Tunguska phenomenon that occurred in Siberia in 1908, where an estimated 15-megaton TNT-equivalent meteoroid detonation caused extensive terrestrial damage, flattening approximately 80 million trees over an area of 2,150 square kilometers [5]. Therefore, it is essential to have a reliable method of predicting the trajectory of these objects to mitigate their potential impact.

With the rise in deep space telescopes and earth-based deep-space surveys, identified rogue objects and Near Earth Asteroids (NEAs) have risen, with over 1.8 Million known asteroids and 25000 known NEAs, with 1100 NEAs identified almost every year [6], [7]. Such a population of rogue objects and the rise in detection of the FMOs and NEAs necessitates their prediction and estimation of their collision trajectory for future evasive maneuvers. Conventional techniques in predicting the trajectory of hazardous asteroids majorly involve comprehensive mathematical modeling that considers various factors, such as the object's size, velocity, and gravitational pull [8]. These models can provide reasonably accurate predictions of the object's trajectory, but their reliance on known parameters limits them. Despite the simulated paths and trajectory modeling, the path of such rogue objects is random to some extent due to thermophysical changing characteristics of the observed vs. model asteroids, non-linear gravitational effects of the sun and planetary position, and errors in observation defining the characteristics to generate the orbital parameters [9], [10]. Accounting for all factors leads to extreme precision in predicting the collision probability. However, lack of fixed pattern and lack of computational efficiency, and high time complexity limits such conventional techniques.

Statistical approximation of asteroids and their orbital parameters and physical characteristics have presented hope for using machine learning (ML) in inter-dependant feature inference and complex calculation towards classification [11]; however, current ML algorithms and techniques also have a high time complexity with limitations in performing complex large scale calculations. Recent technological advances, specifically in quantum machine learning (QML), have opened up new possibilities for predicting and estimating the trajectory of hazardous asteroids [12], [13]. Quantum Machine Learning (QML) involves leveraging quantum computers to perform complex calculations that surpass the computational capabilities of classical computers, as supported by empirical research conducted by Biamonte et al. [14] who demonstrated the potential of quantum algorithms in accelerating pattern recognition, data clustering, and optimization problems, highlighting the advantages of quantum computing over classical methods in the context of machine learning tasks. This technology can potentially revolutionize asteroid hazard prediction and planetary defense by allowing us to consider a more extensive range of factors and make more accurate predictions.

Motivated by the current limitations in conventional techniques for asteroid detection and hazard prediction, we recognize the need for simpler, faster, yet sophisticated approaches. Traditional ML methods have reached a saturation point in efficient classification and prediction tasks, implying a doubt of efficacious asteroid hazard prediction, hence highlighting the potential for QML to fill this gap. To address these challenges, we propose a QML-based approach that employs Variational Quantum Classifiers (VQC) and Pegasos Quantum kernel Support Vector Machines (PQ-SVM) for the prediction and classification of asteroids. In the proposed work, we investigate the application of VQC and PQ-SVM for predicting and classifying potentially hazardous asteroids. VQC, which merges variational quantum circuits with classical optimization methods, forms hybrid quantum-classical classifiers adept at discerning intricate patterns within data [15].

The VQC's capacity for representing high-dimensional data efficiently and its intrinsic robustness against noise render it an attractive candidate for tackling the challenges in asteroid hazard prediction. Conversely, PQ-SVM incorporates the Pegasos algorithm'a primal sub-gradient technique for addressing the SVM optimization problem'alongside quantum kernels [16]. The Fidelity Quantum Kernel in PQ-SVM facilitates the computation of inner products in high-dimensional feature spaces, enabling the separation of complex, non-linearly separable data. The amalgamation of these QML algorithms offers a powerful framework for the classification of hazardous asteroids, bolstering both accuracy and computational efficiency. The QML framework aims to harness the inherent capabilities of quantum computing to enhance the accuracy and efficiency of asteroid hazard prediction. By leveraging VQC and PQ-SVM algorithms, we have the potential to overcome the limitations inherent in classical Machine Learning (ML) methods. VQC algorithm, utilizing the principles of superposition and entanglement in quantum computing, enables efficient exploration of a larger solution space, allowing for more accurate modelling and prediction of possible celestial phenomena. Similarly, PQ-SVM leverages quantum algorithms to enhance classification and regression tasks with potential applications in astronomy and astrophysics, potentially achieving higher accuracy in discerning intricate relationships between variables. These advancements and their inherent utilization, can indirectly help in improving our understanding of the formation and dynamics of the solar system and unveiling the mysteries of the universe, helping reshape our knowledge of

the cosmos. This ultimately contributes to developing more reliable asteroid hazard predictions, allowing us to assess better and mitigate the potential impacts of these celestial objects on our planet.

A. RESEARCH CONTRIBUTION

Major contributions presented in this work have been outlined as follows,

1) We proposed a QML-based classification approach to detect potentially hazardous objects (e.g., asteroids) that pose a risk to our planet. For that, a standard celestial object dataset is utilized that is used by the QML algorithm for prediction purposes. Further, the utilized dataset is converted into a quantum dataset using ZFeatureMap and forwarded to the QML algorithms.
2) The proposed QML-based classification approach adopts different QML algorithms, such as VQC and PQSVC, for binary classification of the potentially hazardous state of an asteroid.
3) Further, we compared the QML algorithms with ML algorithms to showcase the competency of QML algorithms, wherein we utilized different performance parameters, such as accuracy, precision, recall, and F1 score. The results show that QML algorithms outperform ML in terms of accuracy, where PQSVC achieves 98.11% accuracy.

B. ORGANISATION

The following is the organizational structure of this article. Section II discusses the various works in the QML and QC domain and relevant updates on the task domain of asteroid hazard prediction. Section III derives the problem statement for the task and explains the system model. Section IV describes the proposed QML-based methodology for PHA classification, dataset description, pre-processing steps followed, and feature selection steps used for the model. Section V describes and states the experimental setup, different evaluation metrics used, and discussion of results obtained. Finally, section VI wraps up the study's conclusions and comments on the future scope.

II. RELATED WORKS

Over the years, the study of astronomy and space exploration has been instrumental in unraveling the mysteries of the universe. Recent advancements in technology, particularly in quantum machine learning (QML), have opened up new possibilities for predicting the hazardous nature of asteroids and mitigating the potential impact of these objects. However, the study of asteroids and meteorites also highlights their potential threat to our planet, and it is crucial to have reliable methods of predicting their trajectory. Conventional techniques in asteroid hazard prediction have limitations due to their reliance on known parameters. But with the advent of QML, researchers have the potential to improve the accuracy of asteroid hazard prediction significantly. Research in this field has been extensive, with researchers worldwide working to develop better techniques for predicting the hazardous nature of rogue floating asteroids. Table 1 tabulates the comparison of various other state-of-the-art.

In [17], authors have addressed the asteroid-comet hazard by identifying and investigating potential collisions and close approaches of asteroids with the Earth. Resonant returns after encounters with the Earth present a challenge for predicting these events due to the loss of precision in such encounters. The main asteroid of interest in this study is Apophis (99942), for which multiple possible orbits of impacts associated with resonant returns were identified. The study suggests that an early change in Apophis' orbit could avoid these main impacts, which is a feasible solution. Additionally, the study explores potential impacts with Ground asteroid 2015 RN35 and presents 21 possible collisions in this century, including seven collisions with significant gaps reported on the NASA website. Finally, the study presents the observations of three near-Earth asteroids, including 7822 and 68216, which are potentially hazardous, using the ZA-320M telescope at Pulkovo Observatory.

Asteroids such as the Apophis, followed by a case study of Identification and estimation of the S-type Asteroids Itokawa and Bennu, are significant yet limited observed proofs of collision 100% bound to happen in the coming 1000 years. Solely individual identification and trajectory estimation of asteroids that are potentially hazardous for Earth limits other future possibilities; Hence, inferencing other collision possibilities via statistical approximation and pattern analysis are significant and necessary in generating more accurate results with the clustering of asteroid families. A study used supervised-learning hierarchical clustering algorithms to identify family members with high accuracy, precision, and recall values, consistently above 89.5% [18]. In the study, the authors identified six new families and 13 new clumps that appear consistent and homogeneous regarding physical and taxonomic properties. The results highlight the efficiency and speed of machine-learning clustering algorithms for the problem of asteroid family identification.

The hierarchical clustering mechanism results very well in identifying new asteroid families and rubble pile-type rock clumps. The k-Nearest Neighbor algorithm was combined with the Bus-DeMeo taxonomic classification schema to test the ML approach in clustering asteroid families in [19]. The results suggest that the visible wavelength region is more diagnostic of the spectral slope, while the NIR wavelength region is more diagnostic for surface mineralogy. While the author achieved overall accuracy scores (>80%) of the ML test dataset validated the methodology, it fell short of the threshold necessary to replace current methods of classification (>95%). Nonetheless, the study corroborates the overall robustness of the Bus-DeMeo taxonomy in asteroid family clustering.

Various research has also explored the prediction of potentially hazardous asteroids (PHAs) using ML algorithms. In [20], the authors' proposed approach focuses on identifying subgroups of near-Earth asteroids (NEAs) that have a high concentration of PHAs. To achieve this, the author utilized the Support Vector Machines algorithm with RBF kernel to identify the boundaries of PHA subgroups in 2- and 3-dimensional subspaces of orbital parameters. The results show high PHA purity (>90%) and contain >90% of all real and virtual PHAs. These findings can help plan future PHA discovery surveys or asteroid-hunting space missions.

In [21], authors highlight the significance of potentially hazardous objects (PHOs) such as asteroids and comets, that can approach Earth and cause severe damage. The work presents a project aimed at analyzing NASA data on PHOs and predicting whether an Earth-approaching body is hazardous. The authors used the Azure platform for machine learning algorithms to preprocess and model the data, comparing the results of various machine learning models. The Two-Class Decision Forest algorithm achieved the highest accuracy, with a ROC-AUC score of 0.99. However, it is important to note that the dataset used by the authors is of low records, which may limit the generalizability of the results to larger datasets. Nonetheless, the project highlights the potential of machine learning in predicting hazardous PHOs. It is evident that the work underscores the need for continued efforts to monitor and mitigate potential risks from celestial bodies near Earth.

In [22], authors aim to prevent potential asteroid impacts by accurately predicting and classifying Near-Earth Asteroids (NEAs) as potentially hazardous or non-hazardous. The proposed approach uses deep neural networks to learn complex representations of orbital asteroid data and classify them accordingly. By generating an automatic potentially hazardous asteroid detector, the study hopes to speed up the characterization rate of NEAs and contribute to the prevention of natural disasters caused by asteroid impacts. The research highlights the importance of identifying Potentially Hazardous Asteroids early on and classifying them correctly to avoid catastrophic events. The increasing availability of asteroid data through scientific advances provides an opportunity to use ML techniques to tackle this important problem. The field of asteroid hazard prediction has seen significant advancements in recent years, with classical simulations being one of the primary approaches. However, these simulations can be computationally expensive due to the need to predict the collision courses of rogue objects accurately. As a result, using QML has emerged as a promising solution for pattern analysis, generating analytics, and making inferences.

In [23], authors introduce the concept of QML as a dominant paradigm to program gate-based quantum computers in the current NISQ era. QML involves tuning the parameters of a quantum circuit via classical optimization based on data and measurements of circuit outputs. Parameterized quantum circuits (PQCs) can address optimization problems, and generative models and carry out classification and regression. The monograph portrayed in the work by the authors is aimed at engineers with a probability and linear algebra background and provides a self-contained introduction to quantum machine learning. The author has covered the necessary background, concepts, and tools to describe quantum operations and measurements and delves into parameterized quantum circuits, variational quantum eigensolvers, and supervised and unsupervised quantum machine learning formulations. Wherein in the paper, author specifically emphasizes the potential of quantum machine learning in solving complex optimization problems and carrying out advanced inference tasks.

In [24], author discusses the challenge of scalability in QML and proposes a novel framework that utilizes projection-valued measurements (PVM) to expand the output dimension from qubits to 2q. The aim is to achieve multi-class classification by leveraging quantum circuits and the proposed model outperforms state-of-the-art methodologies by 42.2%. The paper also proposes a probability amplitude regularizer to fit the probability distribution of observables to the probability distribution of classification. The authors highlight future work directions, including quantum reinforcement learning with large action spaces or multi-agent settings with fewer qubits and quantum object detection by leveraging PVM for classification and POVM for bounding box prediction. While the results are promising, the paper focuses on datasets with no more than 6 qubits, leaving open the question of how this framework will scale to larger datasets. Additionally, the authors could have discussed the feasibility of implementing their proposed framework on current NISQ quantum computers, as this may pose significant practical challenges. Nonetheless, the proposed framework can potentially address the scalability issue in QML and lead to more complex quantum machine learning applications.

In the field of astronomy, QML has shown promise in various applications such as exoplanet detection and characterization. In [25], the authors demonstrate the application of QML to classify exoplanets using their transit light curves. The study proposes a quantum algorithm that leverages parameterized quantum circuits to perform binary classification tasks. By encoding the light curve data into a feature Hilbert space, the quantum algorithm enables the efficient processing of complex patterns in the data. The study provides insights into the potential of QML for analyzing complex astronomical data, which can be extended to asteroid hazard prediction.

Astrodynamics and celestial mechanics have also seen the application of QML for orbital dynamics and trajectory optimization. Authors in [26] present a quantum algorithm for solving the two-body problem, a fundamental problem in celestial mechanics. The proposed quantum algorithm leverages quantum phase estimation to determine the energy eigenvalues of the two-body Hamiltonian. By reducing the complexity of the problem from exponential to polynomial, the quantum algorithm outperforms classical algorithms in

solving the two-body problem, suggesting that QML can be potentially applied to other problems in astrodynamics, including asteroid hazard prediction.

In another study, [27] authors' have explored the application of QML in space situational awareness (SSA) and space surveillance and tracking (SST) problems. The authors propose a quantum algorithm for detecting and tracking artificial satellites in geostationary orbits. By leveraging quantum amplitude estimation and quantum support vector machines, the algorithm demonstrates improved classification and tracking performance compared to classical methods. This research demonstrates the potential of QML in addressing complex problems in astrodynamics and celestial mechanics and may provide valuable insights for asteroid hazard prediction.

Moreover, in [28], the authors investigate the application of QML for solving the Kepler problem, another fundamental problem in celestial mechanics. The authors develop a quantum variational eigensolver (QVE) to determine the energy eigenvalues and eigenstates of the Kepler Hamiltonian. The QVE algorithm exhibits faster convergence rates and higher accuracy compared to classical variational eigensolvers. The success of the QVE algorithm in solving the Kepler problem highlights the potential of QML for tackling challenging problems in astrodynamics, including asteroid hazard prediction.

Through extensive experimentation and evaluation, QML has proven its ability to outperform current machine learning algorithms and techniques in some domains. Additionally, there is a possibility of expanding QML to other domains in the future. Both supervised and unsupervised learning have shown positive results, indicating that QML can be a dominant paradigm in programming gate-based quantum computers. As the field of QML is still evolving, there is significant potential for further advancements and developments, which can lead to exciting discoveries and applications. Finally, applying QML in the realm of asteroid hazard prediction shows the potential to overcome the limitations associated with traditional simulations. QML algorithms have shown great accuracy, precision, and recall in recognizing asteroid families and PHAs, and can help with taxonomic categorization and forecasting future collisions and near encounters. Further study in this area may yield useful insights and aid in the prevention of catastrophic occurrences caused by asteroid strikes.

A. LIMITATIONS OF CONTEMPORARY ML CLASSIFIERS

Present-day machine learning classifiers, despite their profound predictive abilities, grapple with certain qualitative limitations. Primarily, these classifiers exhibit a strong dependency on both the training dataset's volume and integrity. Insufficient or poor-quality data often lead to an over-generalization phenomenon known as overfitting, wherein the model delineates an excessively complex decision boundary that performs exceedingly well on the training set, but falters on unseen data [29]. While methods like regularization and dropout [30] aims to mitigate overfitting, these techniques offer a partial solution and often necessitate careful hyperparameter tuning.Further, these classifiers typically rely heavily on pre-defined features for model training, making feature engineering a critical task. This process often necessitates domain-specific knowledge, posing a significant limitation, especially in intricate domains [31]. Despite the emergence of deep learning algorithms that facilitate automatic feature extraction, the necessity for manual feature selection isn't entirely negated.

Moreover, these classifiers frequently lack transparency and interpretability, attributing to their characterization as ''black box'' models [32]. This opacity can be especially problematic in domains such as healthcare or finance, where interpretability is as crucial as prediction accuracy. Machine learning classifiers, if trained on biased data, can inadvertently propagate and magnify these biases, leading to potentially skewed or prejudiced outcomes [33]. This latent bias in machine learning models presents serious ethical implications, necessitating more research towards ''fair'' machine learning. Lastly, these models often exhibit a lack of robustness. They tend to be sensitive to minor perturbations in the input data or shifts in the data distribution, resulting in significant fluctuations in predictions [34].

In summary, the limitations of machine learning classifiers, including data dependency and overfitting, reliance on domain knowledge and feature engineering, lack of interpretability, the potential for bias propagation, and stability issues, underscore the importance of ongoing research in improving the qualitative aspects of machine learning models. While modern machine learning classifiers struggle with these drawbacks, the emergence of quantum machine learning (QML) offers a potential remedy. The paradigm shift in processing capabilities brought about by quantum computing, which is based on quantum mechanical principles, may be able to mitigate the drawbacks of traditional machine learning methods.

B. QUANTUM MECHANICS AND COMPUTING

QC is built on the mathematical and physics foundations of quantum mechanics incorporated into the basics of computer science. Using superposition of states and particle entanglement, QC utilizes a novel approach towards information storage via qubit that enables an exponential speedup in computation time, increased parallelism, and complex calculations. Starting with the basic unit of QC, a qubit is a superposition of two basis states $|0\rangle$ or $|1\rangle$ represented by a complex vector as shown in Eq. 1,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1}$$

Here, α and β are complex numbers that satisfy the condition $||\alpha|^2 + |\beta|^2 = 1$ to ensure the qubit is normalized. The coefficients α and β are called probability amplitudes, and they determine the probability of measuring the qubit in the state $|0\rangle$ or $|1\rangle$, which is a translation for a tertiary state of information indicating superposition.

The superposition property of qubits is one of the key features that makes QC so powerful for certain types of computations. It allows a single qubit to perform multiple computations simultaneously, and a quantum system with multiple qubits can exist in an exponentially large number of possible states. To perform multiple complex and parallel computations, quantum computers use sophisticated circuits that involve gates similar to classical computing involving simple binary logic gates. In the view of quantum gates, unitary matrices perform operations on qubits to find the current position of qubits. One important operation is the Hadamard gate, which transforms the basis states as follows,

$$H = \begin{cases} H|0\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle) \\ H|1\rangle = 1/\sqrt{2}(|0\rangle - |1\rangle) \end{cases} \tag{2}$$

The Hadamard gate puts the qubit into an equal superposition of the two basis states. This operation is crucial for many quantum algorithms, including QML, widely used in various QML algorithms. QML algorithms typically involve a combination of quantum and classical components. The quantum component performs specific calculations that are difficult or impossible to perform on classical computers, such as evaluating a quantum kernel function and the classical component performing the final classification or regression step.

III. SYSTEM MODEL AND PROBLEM FORMULATION

The expanse surrounding all celestial bodies, including those with habitable planets like Earth, is commonly characterized as a vacuum. However, this space is not devoid of matter, as it is populated by rogue celestial objects, including meteorites and asteroids, which continuously drift within its bounds. Consider a set of rogue entities, denoted as $\mathcal{A}$, comprising an infinite collection of disparate floating objects, represented by $\{\alpha_1, \alpha_2, \ldots, \alpha_m\} \in \mathcal{A}$. These objects are classified as rogue due to their lack of association with any particular celestial body, rendering them susceptible to collisions amongst themselves. The outcome of such collisions can lead to the formation of even larger objects, which may potentially pose a significant threat to habitable planets such as Earth upon contact.

Deep-space satellites are critical in gathering data such rogue celestial objects, potentially threatening habitable planets like Earth. The primary conventional method for acquiring and gathering useful data about rogue celestial objects' characteristic nature is through satellite imaging via optical and radio astrometry using a set of orbiting satellites $\{\sigma_1, \sigma_2, \ldots, \sigma_n\} \in \mathcal{S}$ in low Earth orbit and deep space orbit. Each captured set of satellite data $\mathcal{I}$ is a partitioned collection of images and radar data $\mathcal{I}_{\alpha_m} = \{\iota_1, \iota_2, \ldots, \iota_k\}$ for $\forall \alpha_m \in \mathcal{A}, \forall \sigma_n \in \mathcal{S}$. Now, all the collected data $\mathcal{I}$ is sent to earth data processing centers $\{c_1, c_2, \ldots, c_j\} \in \mathcal{C}_{earth}$.

$$\mathcal{I} \xrightarrow{f(\theta)} \mathcal{K} \tag{3}$$

In Eq.3, processing operation $f(\theta)$ is a set of analytical astrometrical operations $\{f_1, f_2, \ldots, f_o\} \in f(\theta)$ such as two-body solutions (Sun-Earth-Object), Kepler's laws of planetary motion, radio astronomy, and other solar astronomy-based analytical metric operations applied to extract the needed available data $\{H, diameter, albedo, \ldots, moid\} \in \mathcal{K}$, where H represents absolute magnitude of the rogue object, diameter signifies span of the spherical type structure of object, albedo represents the magnitude of reflected sunlight,moid is the minimum orbit intersection distance, etc. Considering a generalized inference function $g()$, attribute sets $[Diameter, Albedo, a, e, i, tp] \in \mathcal{K}_\varphi \subset \mathcal{K}$ and $[moid, moid\_ld] \in \mathcal{K}_\vartheta \subset \mathcal{K}$ in conjunction with attributed data $\mathcal{K}_{earth}$ about Earth's astrodynamics aids in obtaining the probabilistic possibility of an asteroid's collision course.

$$g(\mathcal{K}_\varphi \cup \mathcal{K}_\vartheta \cup \mathcal{K}_{earth}) \rightarrow \mathcal{P}(Collision) \tag{4}$$

Processing operations applied for the generation of the attributed data are the same equations applied towards the conventional highly accurate simulation of collision for the safety of the planet and bettering planetary defense, i.e., hazard mitigation methods. Simulating the trajectory of a celestial object requires solving a set of complex equations of motion that take into account the object's mass, velocity, gravitational forces, atmospheric drag, and other factors. One commonly used set of equations for simulating the motion of celestial objects is the Keplerian equations $f_1(a, e)$, which describe the motion of an object in an elliptical orbit around a central body.

$$f_1(a,e) = \begin{cases} r = \dfrac{a(1-e^2)}{1+e\cos(\theta)}; & \text{(distance (focus to ellipse))} \\ x = r\cos(\theta+\omega); & \text{(Object's x-coordinate)} \\ y = r\sin(\theta+\omega); & \text{Object's y-coordinate)} \\ z = z_p + r\sin(i)T\sin(\theta+\omega); \end{cases} \tag{5}$$

where, $z_p + r\sin(i)T\sin(\theta+\omega)$ is the object's z-coordinate. Subjecting the obtained attributed data $\mathcal{K}$ to the Keplerian Equations $f_1(a, e)$, we obtain the object's motion characteristics. The above set $f_1(a, e)$ provides either set of coordinates (Polar Coordinates $\mathcal{P}(r, \theta, \omega)$ and Cartesian Coordinates $\mathcal{V}(x, y, z)$) based on available data.

$$\mathcal{P}(r,\theta,\omega) \underset{x,y,z}{\longleftarrow} f_1(a,e) \underset{\theta,\omega}{\longrightarrow} \mathcal{V}(x,y,z) = \vec{r}_t \tag{6}$$

However, the Keplerian equations $f_1(a, e)$ alone may not be sufficient for simulating the motion of rogue celestial objects, especially those on a collision course with Earth. Additional factors such as the object's mass, the Earth's gravitational pull, atmospheric drag, and non-uniform gravity fields must also be considered. One commonly used set of equations for this purpose is the two-body problem $f_2(\vec{r}_t)$, which describes the motion of two objects that interact with each other through gravitational forces at a given moment of time t. The equation for the two-body problem is as follows.

$$f_2(\vec{r}_t) : \frac{d^2\vec{r}_t}{dt^2} = -\frac{GM}{\mathcal{R}^3}\vec{r}_t = \ddot{\vec{r}} \tag{7}$$

TABLE 1. A comparison between the state-of-the-art methodologies and the proposed work.

Author	Year	Objective	Algorithm	Results	Limitation
[35]	2023	Linear Modelling for Quantum Algorithms	Reproducing Kernel Hilbert Space (RKHS)	None	Theoretical Foundation discussion for linearization of data input - import and model
[36]	2023	Classical ML Model comparitive analysis with Project Quantum Kernel Features	Decision Tree, KNN, Naive Bayes and SVC	Avg Accuracy@92%	Reduction in Accuracy for both Original and Relabelled Datasets averaging at 70%
[37]	2022	Generalization Error computation for reduced Dataset Features	QCNN, VAns and GAN	None	No numerical value presentation to make observational inference for making an inference
[38]	2022	Challenging narrative on QC & ML, call for critical debate due to assessment challenges.	None	None	No stated limitation since the basis of the paper does not include quantifiable observation
[22]	2022	Develop a classifier to detect potentially hazardous asteroids automatically.	SVM, Logistic Regression, MLP and KNN	Highest @90% for SVM RBF	The high ceiling of 90% Accuracy for the single algorithm, the rest all have an average accuracy of 60%
[39]	2022	Develop quantum circuit-based classifiers for Image classification with high accuracy.	QNN	@>85% Accuracy in 11 qubits	Classical ML algorithms mapped onto circuits with 11 qubits
[40]	2022	Cluster System exploration with high-quality QC Dataset for QML	Kernel Ridge Regression	MAE of 0.5Kcal/Mol	Training and testing of the model on small clusters of molecules in lower dimensional space
[23]	2022	Tutorial for QML Coding	Discussion of various algorithms	None	None
[21]	2022	Analyze NASA's data on asteroids and use ML to predict hazardous approaches	Decision Tree, Boosted DT, Neural Network, LR, SVM	Average 99% AUC Score	Presentability of the paper does not signify true results
[24]	2022	Develop QML for multi-class classification of images using PVM with improved scalability.	Positive operator- valued measure based QML	∼70% Accuracy	For an image-based dataset with a much larger data receptive field, the accuracy of the model is less compared to other data variations and models
[41]	2021	Review recent advances in quantum classifiers for ML.	None	None	None
[42]	2021	Tutorial for QML Coding	Discussion of various algorithms	-	-

where,

$\overrightarrow{r_t}$; Earth relative position vector
G; Gravitational constant
M; Mass of the Earth
$\mathcal{R} = |r|$; Magnitude of the position vector

The distance in terms of travel time between the Kuiper belt and Earth can be estimated as a 100-120 year period considering the orbital velocities of roaming rogue objects. Hence for plotting a trajectory, it is important to track the object for 100 years to calculate the probabilistic certainty of colliding with Earth. For plotting the trajectory, we need to double integrate the Eq. 7 from time t_i (Initial time of spotting the object) to t_F (Period of 100 years). Therefore the generated set of calculated vector points known as orbit propagation is as follows,

$$\vec{\mathcal{V}} = \iint_{t_i}^{t_F} \ddot{\vec{r}}\, dt^2 \tag{8}$$

Now, each rogue object $\alpha_m \in \mathcal{A}$ has a gravitational effect on each other and is also affected by other planets; hence to compensate for such effects, the N-body solution has to be employed. Now Eq. 7 can be expanded as follows, summing up the gravitational effect on the object α_i

$$f_2(\vec{r}_t) : \frac{d^2\vec{r}}{dt^2} = -G\sum_{i=1}^{n} \frac{m_i}{\mathcal{R}_i^3}\vec{r}_i = \sum_{i=1}^{n} \ddot{\vec{r}}_i \tag{9}$$

Substituting Eq. 9 in Eq. 8 for the integral, we get

$$\vec{\mathcal{V}} = \iint_{t_i}^{t_F} -G\sum_{i=1}^{n} \frac{m_i}{\mathcal{R}_i^3}\vec{r}_i\, dt^2 \tag{10}$$

From Eq. 10 we obtain the propagated orbit of the asteroid accounting for all possible gravitational effects and hence a certainty for its possible collision with earth. However, gravitational compensation doesn't accurately provide probabilistic certainty of possible collision to a certain extent since temporal-based variance can occur, such as inter-object

collision, which may change the course of objects. The position uncertainty of an asteroid is usually relatively small over the time span of the observations, but it usually grows, or stretches, as the object's position is predicted farther and farther into the future. To account for this and to sample such uncertainty, Monte-Carlo Simulations (MC) are performed in which coarse exploration of virtual asteroids ($\beta_1, \beta_2, \beta_3, \ldots, \beta_j \in \mathcal{B}; j = 10000$) is done [43].

$$\mathcal{S}_p = \sum_{j=1}^{10000} \mathcal{MC}(\vec{\mathcal{V}}_j, \beta_j) \tag{11}$$

$$\mathcal{D}_p = \mathcal{F}(\mathcal{S}_p, \mathcal{A}, \mathcal{K}) \tag{12}$$

From the parameter space $\mathcal{U}$ (Orbit Sphere), we obtain a simulated characterized region $\mathcal{S}_p$ (Eq. 11) which is a MC simulation of varied virtual asteroid set $\mathcal{B}$. The above obtained probabilistic distribution of collision uncertainty is filtered further based on the miss-distance at the time of close approach i.e., extended orbit determination filter ($\mathcal{F}$) as shown in Eq. 12 based on the previously available data $\mathcal{K}$ with the incorporation of the simulation object sample (β_j) as a pseudo observation. Final available data on the certainty factor for possible Near Earth Asteroids (NEAs) $\mathcal{N}$ can be depicted as follows,

$$\mathcal{N} : \vec{\mathcal{V}}_{\alpha_i} + \mathcal{D}_{P,\alpha_i}; \quad \forall \alpha_i \in \mathcal{A} \tag{13}$$

Computational efficiency and time complexity are major factors in tracking and predicting the hazardous state of rogue objects. Considering the set of NEAs $\mathcal{N}$, time complexity is in order of years (Eq. 14). The complexity of simulating the motion of a rogue celestial object does depend on various factors such as the accuracy required, the length of the simulation (t_F), the number of objects involved ($[\mathcal{A}, \mathcal{B}]$), and the computational resources available. For example, simulating the motion of an object with high accuracy over a long period may require a large number of iterations (x) and significant amounts of memory storage, making it computationally expensive.

$$\mathcal{T}(g(\mathcal{A}, \mathcal{MC}(\vec{\mathcal{V}}, \mathcal{B}_j), x) \xrightarrow{Sentry} \mathcal{N}) = n \cdot Years \tag{14}$$

In contrast, QML algorithms and quantum computation techniques can efficiently flag hazardous rogue objects, wherein the QML algorithms can be used to process large amounts of data and identify patterns that are difficult or impossible for classical algorithms to detect. Subjecting the processed data $\mathcal{K}$ to QML algorithms $\{\rho_1, \rho_2, \rho_3, \ldots, \rho_p\} \in \mathcal{Q}$, the subjective time complexity is reduced parallel to efficiently training a highly accurate model for flagging the sample objects of the collected dataset. The main objective for the problem of efficient and accurate hazard prediction can be defined as a set of multiple maximizing and minimizing functions. In this study, 2 major QML algorithms, namely, VQC, and PQSVC have been employed wherein the best-performing algorithm is PQ-SVC. Hence the objective function $\mathcal{O}$ for algorithm PQ-SVC ρ_1 can be defined as minimizing hinge loss function $h()$ for predicted (ψ) and true label (μ) delta (δ) (15).

$$\mathcal{O}(\rho_1, \mathcal{K}) \rightarrow \min_{\kappa \in \mathcal{K}} h(k, \rho_1, \delta(\mu, \psi)) \tag{15}$$

IV. PROPOSED METHODOLOGY

In this proposed methodology (as shown in FIGURE 1), a preprocessed and cleaned dataset is used for analysis using quantum algorithms. Exploratory Data Analysis (EDA) techniques are used to gain insights into the dataset and identify any patterns or trends that can be used to improve the accuracy of the predictions. Two quantum algorithms, the VQC algorithm and the PQSVC algorithm, are used to classify and predict the potential hazard of asteroids and meteorites. The proposed methodology shows promising results and has the potential to improve the accuracy and speed of potential asteroid hazard prediction.

A. DATA COLLECTION AND DATASET DESCRIPTION

For rogue entities $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathcal{A}$, collected observed data I is processed via processing centers $\mathcal{C}$. Processed data $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathcal{K}, \forall \alpha_m \in \mathcal{A}$ is stored in a repository maintained and updated by NASA Jet Propulsion Laboratory (JPL). The dataset utilized in developing this research is sourced from Kaggle [44], wherein the authors Hossain and Zabed have generated the dataset from a maintained repository of NASA JPL. The dataset contains 958524 records of deep space objects described across 45 attributes with two target columns of Near Earth Object Flag and Potential Hazard Flag. Only the potential hazard flag column has been considered the target column for this research. In the target column, only 2066 records of asteroids are flagged as potentially hazardous, and the rest, 956458, are flagged as non-hazardous. Due to the general Quantum Computing Constraints, 4132 records of rogue objects have been included in the final dataset wherein there are original 2066 Hazardous objects and 2066 random sampled subsets of the 956458 non-hazardous rogue objects. The collection $\mathcal{K}$ details the rogue object's orbital characteristics (e.g., semi-major axis, eccentricity, perihelion distance, inclination, etc.), physical characteristics, and identification-naming attributes. It is pre-processed and cleaned to ensure the dataset is acceptable for quantum algorithms analysis.

B. DATASET PREPROCESSING

The preprocessing of the dataset involves several steps to ensure that the data is suitable for analysis using quantum algorithms. The dataset, denoted by $\mathcal{K}$, is obtained from the JPL repository through a dataset filter provided by JPL. The publishing author [44] utilizes this filter to obtain the filtered dataset, denoted by $\mathcal{K}_n$ where $n = 958524$ records. This filtered dataset is stored in a comma-separated value (CSV) file, which is loaded into a data frame using the pandas library

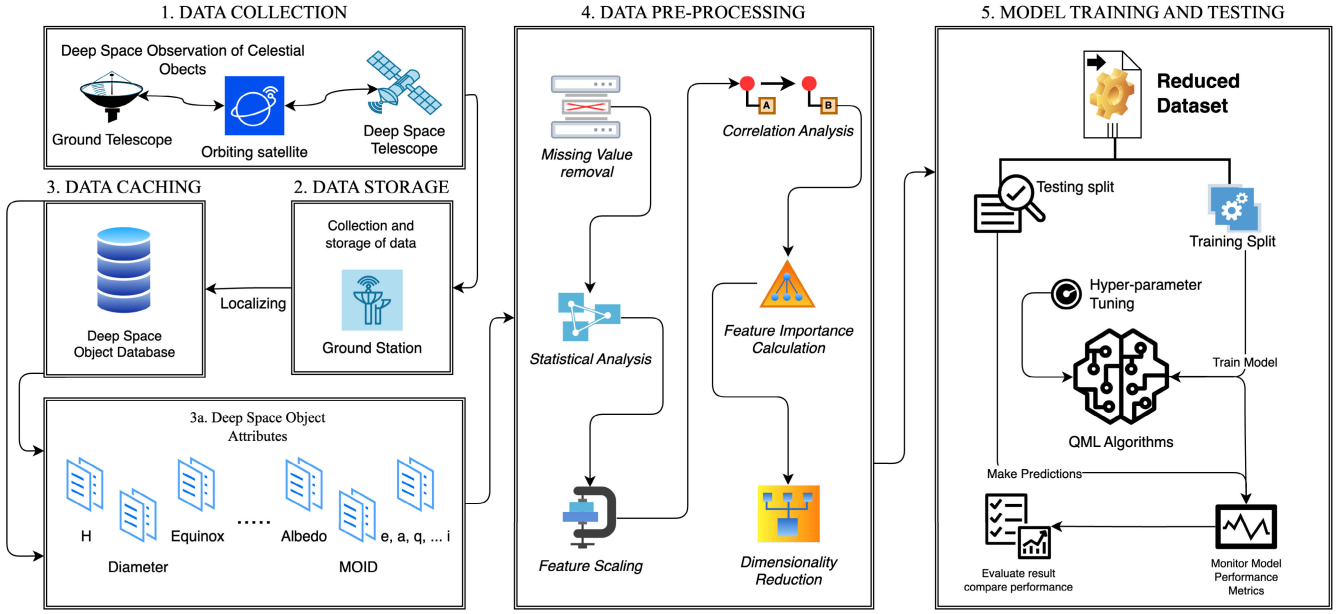


FIGURE 1. Proposed methodology for QML-based for potentially hazardous asteroid classification.

(see Eq. 16).

$$\mathcal{K}_n \xrightarrow[Pandas]{csv-conversion} \mathcal{D}_f \tag{16}$$

All the preprocessing operations are performed on the converted dataset pandas dataframe. After the dataset has been loaded onto the pandas dataframe, the dataset is searched for 'Not a Number' (NaN) type values, obtaining a set of columns having the NaN values. Post obtaining the column name, the column set is segregated based on the columns' characteristics, i.e., one with negligible relevance is eliminated, and the rest are retained for further processing (Eq. 17).

$$\mathcal{D}_f \xrightarrow[Search]{NaN} \mathcal{L}_{NaN} \xrightarrow[segregation]{Importance} \mathcal{L}_p(\text{partitioned set}) \tag{17}$$

In Eq. 17, partitioned set $\mathcal{L}_p = [\mathcal{L}_e$ (columns for removal) $| \mathcal{L}_r$ (columns for retention) $]$. From the dataframe $\mathcal{D}_f$, Columns $\mathcal{L}_e$ are removed. The retained columns $\mathcal{L}_r$ are yet to be processed; however, before processing them, the dataset is searched for columns with object type datatype, i.e., string type datatype (Eq. 18).

$$\mathcal{D}_f - \mathcal{D}_f[\mathcal{L}_r] \xrightarrow{reduced} \mathcal{D}_r \xrightarrow[filter]{object-type} \mathcal{L}_o \tag{18}$$

The column set $\mathcal{L}_o$ containing the name of object type columns obtained from Eq. 18 is used to process dataframe $\mathcal{D}_r$, i.e., label encode and replace the NaN values by mean value using SimpleImputer (Eq. 19).

$$\mathcal{D}_r[\mathcal{L}_o] \xrightarrow[encoder]{Label} \mathcal{D}_l \xrightarrow[imputer]{Simple} \mathcal{D}_i \tag{19}$$

We obtain a cleaned dataset from Eq. 19 with all unnecessary columns removed and NaN columns processed for further steps. Statistical analysis is conducted where skewness analysis, $(\mathcal{S}_k(D))$, is performed on all columns. Further, the necessary normalization technique is applied based on the skew nature of the value distribution in each column.

$$\mathcal{S}_k(\mathcal{D}_i) : \begin{cases} \mathcal{S}_k(\mathcal{D}_i) > 0.5 : \mathcal{B}_c(D) \\ \mathcal{S}_k(\mathcal{D}_i) \in (-0.5, 0.5) : \mathcal{Z}_s(D) \\ \mathcal{S}_k(\mathcal{D}_i) < -0.5 : \mathcal{G}_l(D) \end{cases} \tag{20}$$

In Eq. 20 $\mathcal{B}_c(D)$, $\mathcal{Z}_s(D)$ and $\mathcal{G}_l(D)$ denote the normalization techniques BoxCox power transformation, Z-Score normalization and Log transformation applied based on the respective skew value.

$$\mathcal{B}_c(\mathcal{D}_i) \cup \mathcal{Z}_s(\mathcal{D}_i) \cup \mathcal{G}_l(\mathcal{D}_i) \xrightarrow[dataset]{normalized} \mathcal{D}_n \tag{21}$$

After normalizing the dataset using Eq. 21, we analyzed that there is a significant imbalance in the frequency of records for each of the binary classes ($\mathcal{C}_0$: $Non-Hazardous, \mathcal{C}_1$: $Hazardous$), where one class ($\mathcal{C}_0 = 956458$) has a substantially higher frequency than the other ($\mathcal{C}_1 = 2066$), a sampling process $\mathcal{S}_d$ is performed. The records from the higher frequency class are sampled (random sampling) equally to match the number of records in the lower frequency class. Hence the final subset of the dataset contains a total of 4132 records subject to dimensionality reduction. Algorithm 1 shows all the steps used in data preprocessing.

$$\mathcal{D}_n : \begin{cases} freq(\mathcal{D}_n[\mathcal{C}_1]) = n \\ \mathcal{D}_n[\mathcal{C}_0] = \mathcal{S}_d(n, \mathcal{D}_n[\mathcal{C}_0]) \end{cases} \xrightarrow[balancing]{sampling} \mathcal{D}_b \tag{22}$$

QML requires manipulating quantum states, which can be challenging when dealing with high-dimensional data. This is because the number of quantum resources required to

represent and manipulate the data increases exponentially with the dimensionality of the data. Therefore, to reduce the dimensionality of the data, classical ML algorithms are utilized before applying QML techniques, helping reduce the number of quantum resources required and making the problem more tractable for quantum computers. Furthermore, reducing the dimensionality of the data can also help mitigate the effects of noise and other errors inherent in quantum systems, improving the accuracy of the QML algorithms.

Post-normalization and dataset balancing (Eq. 22), the dataset $\mathcal{D}_b$ is subject to a series of classical machine learning algorithms for calculating feature importance $\mathcal{F}_e(m, \mathcal{D})$. These algorithms are employed to reduce the dimensionality of the data, thereby easing the computational requirements of the subsequent Quantum Machine Learning (QML) model. The suite of nine selected algorithms $m_1, m_2, m_3, \ldots, m_e \in \mathcal{M}$, where $e = 9$, include the ExtraTrees classifier, RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier, DecisionTreeClassifier, IsolationForest, VarianceThreshold, SelectKBest, and RandomForestEnsemble. Ensemble methods, ExtraTrees and RandomForestClassifiers, manage high-dimensional data by employing multiple decision trees. Their output is averaged, providing robust feature importance rankings and reducing overfitting [45], [46].

Similarly, AdaBoostClassifier and GradientBoostingClassifier are ensemble methods that function by sequentially fitting models and adjusting weights of incorrectly predicted instances, hence focusing on "harder-to-predict" instances and enhancing overall model performance [47], [48]. The DecisionTreeClassifier, a fundamental algorithm, ranks feature importance based on how features are used to split data within the tree, offering interpretability [49]. The IsolationForest algorithm, distinct in its approach, isolates anomalies or outliers in the dataset by randomly selecting a feature and a split value. This method provides an understanding of which features contribute most to these anomalies [50]. Variance threshold is a straightforward baseline feature selection method that discards features with low variance, presuming they contain less useful information [31]. Finally, SelectKBest, a univariate feature selection method, applies statistical tests to select the 'k' best features. Here, it is used to identify the top 4 most important features [51].

$$\bigcup_{i=1}^{e} \mathcal{F}_e(m_i, \mathcal{D}_b) = \mathcal{F}_a \xrightarrow[voting]{frequency} \mathcal{T}_{f4} \xrightarrow{final} \mathcal{D}_F = \mathcal{D}_b[\mathcal{T}_{f4}] \tag{23}$$

Therefore, using the aforesaid algorithms, we determine the feature importance from each method, creating the feature importance set $\mathcal{F}_a$ for the normalized dataset (Eq. 23). By ranking these features, we apply a frequency voting approach to further refine the selection. In this voting process, we create two sets of highly important features. The first set includes features that have been deemed important by more than four algorithms, and the second set includes those selected by more than five algorithms. The intersection of these sets will contain the most consistently significant features across the different methods, further validating their importance. Subsequently, from these intersection sets, we select the top 4 most important features, denoted as $\mathcal{T}_{f4}$. These are deemed the most significant and will be used to reduce the dimensionality of the dataset. The final dataset containing these top 4 features is generated, which is then utilized for training and testing the Quantum Machine Learning (QML) model. This approach, which combines multiple classical machine learning methods for feature selection, helps to manage high-dimensional data, reduces the quantum resources required, and mitigates the noise and errors inherent in quantum systems.

Algorithm 1 Dataset Preprocessing for Quantum Algorithms

Inputs: Filtered dataset $\mathcal{K}_n$ **Output:** Preprocessed dataset $\mathcal{D}_F$ for QML training and testing

procedure Preprocess($\mathcal{K}_n$)
  $\mathcal{D}_f \leftarrow$ Pandas_csv_conversion($\mathcal{K}_n$) (Eq. 16)
  $\mathcal{L}_p \leftarrow$ NaN_segregation($\mathcal{D}_f$) (Eq. 17)
  $\mathcal{D}_r, \mathcal{L}_o \leftarrow$ Remove_Reduce_Columns($\mathcal{D}_f, \mathcal{L}_p$) (Eq. 18)
  $\mathcal{D}_i \leftarrow$ Label_Encode_Impute($\mathcal{D}_r, \mathcal{L}_o$) (Eq. 19)
  $\mathcal{D}_n \leftarrow$ Normalize($\mathcal{D}_i$) (Eq. 20, 21)
  $\mathcal{D}_b \leftarrow$ Balance_Sampling($\mathcal{D}n$) (Eq. 22)
  $\mathcal{T}_{f4} \leftarrow$ Feature_Importance_Selection($\mathcal{D}_b$) (Eq. 23)
  return $\mathcal{D}_F$
end procedure

Algorithm 2 Procedure for QML Classifiers With Ansatz Variations for Asteroid Hazard Prediction

Inputs: Filtered dataset of Rogue Objects, denoted by $\mathcal{K}n$ **Output:** Classification of Hazardous Nature of the Asteroid - 1(Hazardous) / 0 (Non-Hazardous)

procedure QML(D)
  $\mathcal{D}F \leftarrow$ Preprocess($\mathcal{K}n$) (Algorithm 1)
  $train_d, test_d$ = train_test_split($\mathcal{D}_F$, test_size=0.2)
  n_qubits = 4
  quantum_instance ← QuantumInstance(backend)
  feature_map ← ZFeatureMap(n_qubits)
  fidelity_quantum_kernel ← FidelityQuantumKernel(feature_map)
  for ansatz in [RealAmplitudes(), EfficientSU2()] **do**
    $M1 \leftarrow$ VQC(ansatz, feature_map)
    $M1$.train($train_d$)
    $p_1 \leftarrow M_1$.predict($test_d$)
    evaluate_performance($test_d, p_1$)
  end for
  for C in [0.01, 0.1, 1, 1000] **do**
    $M_2 \leftarrow$ PegasosQSVM(fidelity_quantum_kernel, C=C)
    M_2.train($train_d$)
    $p_2 \leftarrow M_2$.predict($test_d$)
    evaluate_performance($test_d, p_2$)
  end for
  compare_performance(M_1, M_2)
end procedure

C. QUANTUM MACHINE LEARNING

In this work, two algorithms, namely PQSVC and VQC have been trained and tested for quantum modeling toward potential asteroid hazard prediction.

1) VARIATIONAL QUANTUM CLASSIFIER (VQC)

The variational quantum classifier takes a variational approach toward quantum circuit training for input data classification. Being a hybrid, this classical quantum classifier has three key components: a primary circuit, a variational circuit, and a classical optimizer. The complete procedure of training a variational quantum classifier is divided into 4 steps: input encoding, variational quantum circuit, measurement encoding, and a classical optimizer for performance optimization. Initially, the input data is encoded into a quantum state using a mapping function. One common method is the ZFeatureMap, which is defined as.

$$ZFeatureMap(x) = \prod_{i=1}^{n} Z(\phi_i)x_i \tag{24}$$

where, x is the input data, n is the number of features, $Z(\phi_i)$ is the Pauli-Z gate with a rotation angle of ϕ_i, and x_i is the i^{th} feature of the input data. The encoded quantum state can be expressed as.

$$|\psi_{in}\rangle = ZFeatureMap(x)|0\rangle \tag{25}$$

where $0\rangle$ is the initial state of the qubits.

$$|\psi_{in}\rangle = \sum_{x_i \in 0,1} \sqrt{p(x_i)} \cdot |x_i\rangle \tag{26}$$

where, $p(x_i)$ is the probability of measuring the i^{th} qubit in the computational basis state $x_i\rangle$. After the input data is encoded into the quantum state, it is passed through a variational quantum circuit, which consists of a series of parameterized gates. The output state of the circuit is then measured to obtain a classical bit string. Two variations of the parameterized circuit employed are the RealAmplitudes circuit and the EfficientSU2 circuit. The RealAmplitudes circuit is a simple circuit that consists of alternating layers of single-qubit rotations and entangling gates. The circuit parameters are the rotation angles and the entangling gate types. The RealAmplitudes circuit can be expressed as.

$$U(\boldsymbol{\theta}) = e^{-i\theta_p H_p} e^{-i\theta_{p-1} H_{p-1}} \cdots e^{-i\theta_1 H_1} \tag{27}$$

where, $\boldsymbol{\theta}$ is the set of variational circuit parameters, H_i is the Hamiltonian for the i^{th} layer of the circuit, and θ_i are the rotation angles for the i^{th} layer. FIGURE 2 illustrates the trainable parameterized circuit used for data encoding. The EfficientSU2 circuit is a more complex circuit that consists of layers of single-qubit rotations and entangling gates, similar to the RealAmplitudes circuit. However, the EfficientSU2 circuit uses a more efficient parameterization for the single-qubit rotations, which reduces the number of

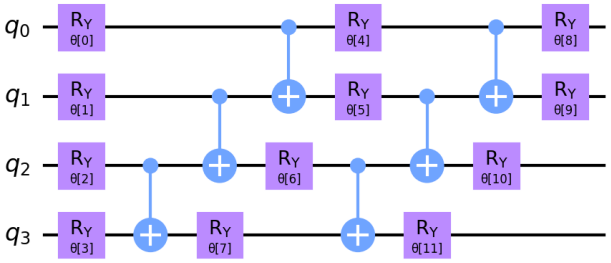


FIGURE 2. Parameterised circuit.

circuit parameters. The circuit can be expressed as.

$$U(\boldsymbol{\theta}) = \prod_{j=1}^{m} U_{\text{ent}}(\boldsymbol{\theta} j) U_{\text{single}}(\boldsymbol{\phi}_j) \tag{28}$$

where, $\boldsymbol{\theta}_j$ and $\boldsymbol{\phi}_j$ are the sets of entangling and single-qubit rotation parameters, respectively, for the j^{th} layer of the circuit, and m is the number of layers. The output state of either of the two circuits can be written as.

$$|\psi_{out}\rangle = U(\boldsymbol{\theta})|\psi_{in}\rangle \tag{29}$$

The measured bit string for either of the two parameterized circuit output states is represented as a classical probability distribution.

$$P(z|\boldsymbol{\theta}) = |\langle z|U(\boldsymbol{\theta})|\psi_{in}\rangle|^2 \tag{30}$$

where z is a bit string representing the measurement outcome. The classical bit string obtained from the circuit output is then mapped back to a quantum state using measurement encoding. This is done by applying a set of Pauli-X gates to each qubit corresponding to a bit in the classical bit string and then applying a phase flip gate to the resulting state. The measurement encoding can be represented mathematically as.

$$|\psi_{me}\rangle = \prod_{j=1}^{n} (-1)^{z_j} X_j |\psi_{out}\rangle \tag{31}$$

where n is the number of qubits in the circuit, z_j is the j^{th} bit of the classical bit string, and X_j is the Pauli-X gate acting on the j^{th} qubit. The final quantum state is used to compute the cost function, which measures the error between the predicted and actual classification outputs. One common cost function is the binary cross-entropy function, which is defined as.

$$C(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(P(y_i = 1|\boldsymbol{\theta}, x_i)) \tag{32}$$

$$+ (1 - y_i) \log(P(y_i = 0|\boldsymbol{\theta}, x_i))] \tag{33}$$

where, $\boldsymbol{\theta}$ is the set of variational circuit parameters, N is the number of input samples, x_i is the i^{th} input sample, y_i is the corresponding target output, and $P(y_i = 1|\boldsymbol{\theta}, x_i)$ is the probability of predicting a positive output for the i^{th} input sample, given the variational circuit parameters. In summary, the input encoding for a VQC using the ZFeatureMap

involves applying the Pauli-Z gates to the input data, and the VQC itself involves using a variational quantum circuit with either the RealAmplitudes circuit or the EfficientSU2 circuit, where the circuit parameters are optimized to minimize the loss function.

2) PEGASOS QUANTUM SUPPORT VECTOR CLASSIFIER

The PQSVC is a QML algorithm that accelerates the computation of support vector machines (SVMs) using quantum mechanics. It is based on the classical Pegasos algorithm, a variant of the stochastic gradient descent (SGD) algorithm commonly used to optimize SVMs. The Pegasos algorithm is highly efficient, easy to implement, and has excellent convergence properties. PQ-SVC builds upon this foundation, using quantum computing to solve large-scale SVM problems more efficiently than classical algorithms. The training process of PQ-SVC is a non-linear process. This quantum version of the Pegasos algorithm is the basis of the PQ-SVM algorithm, which is formulated as follows:

$$\mathcal{O}_{pegasos} = \min_{\boldsymbol{w}} \frac{1}{2}|\boldsymbol{w}|_2^2 + C\sum_{i=1}^{m} \max(0, 1 - y_i \boldsymbol{w}^T \boldsymbol{x}_i)^2 \quad (34)$$

where $\boldsymbol{w}$ is the weight vector, $\boldsymbol{x}_i$ is the input vector, y_i is the corresponding label, and C is a regularization parameter. The first term in the objective function $\mathcal{O}_{pegasos}$ represents the regularization term, while the second term represents the hinge loss function. The optimization problem aims to find the weight vector that minimizes the objective function. The PQ-SVM algorithm uses a quantum feature map $\Phi(\boldsymbol{x})$ to map the input vectors into a high-dimensional quantum state, which is then used to compute the inner product between the weight vector and the quantum state. The quantum feature map is defined as follows.

$$\Phi(\boldsymbol{x}) = \frac{2}{\sqrt{d}} \sum_{j=1}^{d} \sin\left(\frac{\pi j}{d}\right) |j\rangle |x_j\rangle \quad (35)$$

where $\boldsymbol{x}$ is the input vector of dimension d, and $|j\rangle$ is the j-th computational basis state in a quantum system. Prior to mapping the input vector to the quantum state, classical data is converted to a quantum basis state via Pauli Z-evolution circuits. FIGURE 3 shows the circuit utilized in encoding the classical data to qubit basis state.

The Fidelity Quantum Kernel is used to compute the kernel matrix in the PQ-SVM algorithm. The kernel matrix is a matrix of inner products between the feature maps of all pairs of input vectors, which is used to calculate the decision boundary between the classes. The Fidelity Quantum Kernel is defined as follows:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = |\langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle|^2 \quad (36)$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are two input vectors, and $\Phi(\boldsymbol{x})$ is the quantum feature map of input vector $\boldsymbol{x}$. This kernel is based on the fidelity of the quantum states represented by the feature maps of the input vectors. It is sometimes called the Fidelity Kernel because it measures the overlap between two quantum states

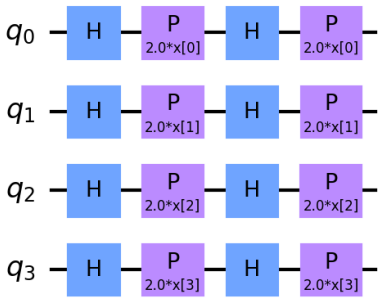


FIGURE 3. ZFeature map for classical to qubit encoding.

regarding their fidelity. Once the kernel matrix is computed, the PQ-SVM algorithm uses it to find the optimal weight vector that separates the classes with a maximum margin by minimizing the regularized empirical loss function. The weight vector is updated using the classical stochastic gradient descent (SGD) algorithm, with each step involving a projection onto the ball of radius $\sqrt{\frac{2}{\lambda}}$, where λ is the regularization parameter. The update rule is given by:

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \eta_t \left(\lambda_t \boldsymbol{w}_t + \frac{1}{mC} \sum_{i=1}^{m} [y_i > [\boldsymbol{w}_t \cdot \Phi(\boldsymbol{x}_i)]_+] \Phi(\boldsymbol{x}_i) \right) \quad (37)$$

where $\boldsymbol{w}_t$ is the weight vector at iteration t, η_t is the learning rate at iteration t, $\lambda_t = \frac{1}{t}$ is the regularization parameter at iteration t, m is the batch size, C is the regularization parameter, y_i is the label of input vector $\boldsymbol{x}i$, and $[\cdot]+$ denotes the hinge loss function. The hinge loss function is defined as follows.

$$[\boldsymbol{w} \cdot \Phi(\boldsymbol{x})]_+ = \max(0, 1 - y\boldsymbol{w} \cdot \Phi(\boldsymbol{x})) \quad (38)$$

where, y is the label of the input vector. The PQ-SVC algorithm is a quantum version of the Pegasos algorithm used to optimize SVMs. The algorithm aims to minimize the hinge loss function in a binary classification problem, with the weight vector as the decision boundary. PQ-SVM, on the other hand, is a quadratic program that finds the optimal hyperplane separating the classes with a maximum margin. The PQ-SVM algorithm uses a quantum feature map to map the input vectors into a high-dimensional quantum state and the Fidelity Quantum Kernel to compute the kernel matrix. The algorithm iteratively applies the quantum feature map, kernel matrix computation, and classical SGD optimization until convergence.

The regularization parameter C in the PQ-SVM algorithm fundamentally influences the balance between the regularization term and the misclassification term in the SGD update rule. It inversely scales the aggregated hinge loss term in the SGD update rule, thus directly governing the weightage of misclassified points in the gradient update of the weight vector $\boldsymbol{w}_t$. A larger value of C effectively decreases the contribution of the hinge loss term in the SGD update, making the

algorithm more tolerant of individual misclassifications. This typically results in a classifier with wider margins, promoting better generalization of unseen data. However, it simultaneously increases the risk of higher bias or underfitting, as the model may not capture all patterns in the training data. Conversely, a smaller value of C increases the influence of the misclassified points in the weight vector update. This urges the algorithm to fit these points more strictly, leading to a classifier with narrower margins that fits the training data more closely. While this can capture more complex patterns in the data, it also increases susceptibility to overfitting, especially in the presence of outliers or noisy data.

Our exploration started with a lower value of $C = 1$, corresponding to a simpler model with a larger margin and a higher tolerance for misclassifications. As we increased C to 10, the margin reduced and the model's complexity increased in an attempt to more accurately fit the data. However, our results indicated a slight decrease in model performance compared to $C = 1$. Subsequently, we tested the model with $C = 100$, further pushing toward a more complex model. Interestingly, this also led to inferior performance compared to $C = 1$. This underlines the fact that simply increasing the model complexity doesn't necessarily lead to improved performance, as it may cause the model to overfit the training data. Driven by these insights, we decided to test the opposite end of the spectrum by setting $C = 1000$. This choice corresponds to a much more complex model with a smaller margin and fewer tolerated misclassifications. Notably, this model outperformed all previous settings on our dataset, demonstrating that in our specific case, a more complex model with stricter classification constraints was more suitable.

Therefore, the parameter C has a critical role in determining the behavior of the PQ-SVM algorithm, striking a balance between bias and variance in the model. Careful tuning of C is essential to achieve the optimal performance of the PQ-SVM model. The regularization parameter and the learning rate control the convergence speed and generalization capability of the algorithm. A predefined tolerance level determines the algorithm's convergence, which measures the difference between the current and previous weight vectors. The training process of PQ-SVC is a non-linear form of tracked loss updation with an increase in complexity at every step.

V. RESULTS AND ANALYSIS

A. EXPERIMENTAL SETUP AND SIMULATION ANALYSIS

The proposed architecture was simulated on the IBM Quantum Experience cloud computing facility, utilizing quantum resources that included a context-aware 32 qubit QASM simulator and 7 qubit free access quantum computers located across 3 different locations. These quantum resources provided the necessary computational power to carry out the simulations and experiments. To thoroughly assess the algorithms' performance, multiple runs were scheduled and conducted. This involved testing the algorithms' handling and efficiency with dataset size variations and evaluating their performance on different quantum systems. After running multiple experiments, it was concluded that current NISQs cannot be utilized to analyze QML algorithms for large datasets with higher dimensions. Hence, currently, NISQs can be operated up to 7 qubits without noise and error with less computational time with roughly dataset size being in the range of 5000 to 10000 records.

During the simulation analysis, careful consideration was given to the limitations and unique characteristics of each quantum system. This ensured that the algorithms' behavior in real-world scenarios was effectively evaluated, taking into account the specific attributes of the quantum resources used. To support the simulation analysis, a range of Qiskit libraries were employed. These libraries, including qiskit-aer, qiskit-ibm-experiment, qiskit-ibm-provider, qiskit-ibm-runtime, and qiskit-machine-learning, played a crucial role in facilitating quantum circuit simulation, executing experiments on IBM Quantum systems, interacting with the quantum computers, and constructing and training the Quantum Machine Learning (QML) model. The implementation code was done in Python on the IBM Quantum Experience iPython Environment. The results for the classical ML comparison were carried out in MATLAB.

Through the comprehensive evaluation of the algorithms on various quantum systems and the use of dedicated quantum resources, the proposed architecture's effectiveness, efficiency, and robustness were rigorously examined. This thorough analysis serves to validate the suitability and potential of the proposed approach for the given task. Table 2 displays the simulation parameters used in the proposed methodology.

B. EVALUATION METRICS

Evaluation metrics enable quick comparison of models and comprehensive assessment of their performance on training and new data. This ensures well-trained models with effective generalization, making evaluation metrics a vital tool for ensuring model efficacy and reliability, especially in fields where a detailed understanding of model performance is necessary. The ensuing discussion concerns a range of metrics that have been utilized in selecting the final QML model from among multiple algorithms, including variations based on hyper-parameters. The problem task's target is binary classification; therefore, the metrics for analyzing the model performance are accuracy, f1_Score, precision, recall, and confusion matrix, described as follows.

1) CONFUSION MATRIX

The confusion Matrix summarizes the complete characteristics of a model's performance. Consisting of Four elements as below, a combination of them indicates the model's performance bias. The elements of the confusion matrix are as follows,

- ϵ (True Positives): The number of correctly identified positive instances by the classifier.
- ζ (False Positives): The number of incorrectly identified positive instances by the classifier.

TABLE 2. Simulation parameters of the proposed methodology.

Parameters	Values
Quantum computer parameters	
IBM Cloud Service	Mail ID
Quantum Computer	ibmq_qasm_simulator
Runtime Service	ibm_quantum
Backend Provider	qasm
General Parameters	
Qubits	32 (In use = Number of Features)
Feature Map	First Order Pauli-Z Evolution Circuit
Number of Features	4 (Voted Selected Features)
VQC Algorithm Parameters	
Kernel	QuantumKernel
Optimizer	COBYLA
Ansatz 1	RealAmplitudes
Ansatz 2	EfficientSU2
Sampler	Base Class Sampler
Iterations	75
Reps	3
Pegasos SVM	
C	0.01, 0.1, 1, 10, 100 and 1000
Kernel	FidelityQuantumKernel
Reps	1
tau	100

- η (True Negatives): The number of correctly identified negative instances by the classifier.
- θ (False Negatives): The number of incorrectly identified negative instances by the classifier.

A comparative display of the confusion matrix for the various algorithms has been shown in FIGURE 4. Prediction of Asteroid being Hazardous is the primary target, hence it is the '1' label in the model. Based on the values of the confusion matrix, three differences are present across the algorithms Employed. Model Characteristics and varying training times lead to,

- Higher Sensitivity ($\epsilon > \eta$)
- Higher Specificity ($\epsilon < \eta$)
- Balanced Classification ($\epsilon \approx \eta$)

Sensitivity and specificity are two essential metrics that directly measure a model's ability to classify positive and negative instances correctly. Sensitivity, also known as the true positive rate, represents the proportion of true positive instances correctly identified by the model. Specificity, also known as the true negative rate, indicates the proportion of true negative instances accurately classified by the model. Given that detection of hazardous asteroids is as important as the low expectancy of an asteroid being life-threatening, therefore it is required that the model have equal sensitivity and specificity towards the PHA detection. As observed in the confusion matrices of the algorithms in FIGURE 4, VQC with EfficientSU2 as the ansatz or trainable parameterised circuit has a balanced specificity @397 and sensitivity @389 with Low Type I error @28 and even lower Type II error @13. Also algorithm Pegasos QSVC with C = 1000 has balanced specificity @408 and sensitivity @402, both errors extremely low with Type-I error @5 and Type-II error @12.

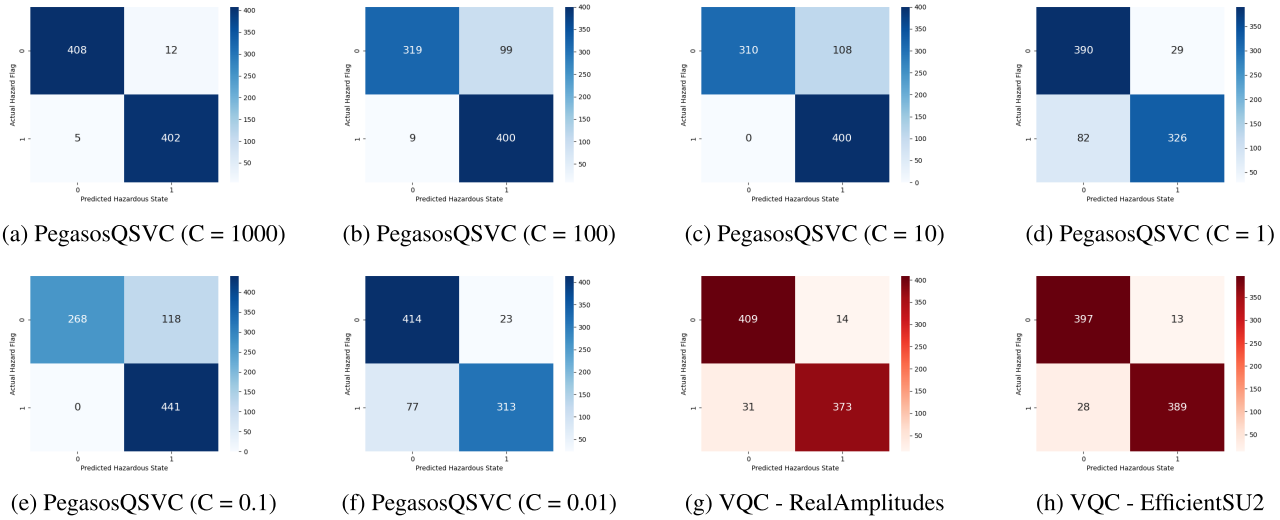
Algorithms VQC with RealAmplitudes ansatz, PegasosQSVC with C = 100, C = 10, C = 1 and C = 0.01 have a higher specificity. VQC with RealAmplitudes ansatz has a higher specificity @409 with sensitivity @373, Low Type I error @31 and even lower Type II error @14. The PegsosQSVC algorithm with C = 100 and C = 10 have an equal specificity @400 with low sensitivity @319 and @310, respectively, with low Type I error @9 for C = 100 and 0 Type I error for C = 10, while the Type II error for both algorithms is in the same range with C = 100 having Type II error instances @99 and C = 10 having Type II error instances @108. Similarly, the PegsosQSVC algorithm with C = 1 has a higher Specificity @390 with low sensitivity @326, and higher Type I error instances @82 than the Type II error instances @29. Also, The PegsosQSVC algorithm with C = 0.01 has a higher Specificity @414 with low sensitivity @313 and higher Type I error instances @77 than the Type II error instances @23. Further, for all c variations for the PegasosQSVC classifier, C = 0.1 is the worst performer with significantly high sensitivity @441, lower specificity @268, and 0 Type I error; however, much higher Type II error instances @118. Algorithm 2 shows the procedure for QML classifier with ansatz various for the proposed work.

2) PRECISION, RECALL AND F1-SCORE

Precision, recall, and F1-score are essential evaluation metrics in the field of machine learning, particularly for classification problems. These metrics provide valuable insights into the performance of a classification model. Precision measures how accurately a model can predict positive instances, while recall measures how well it can identify positive instances. F1-score is a balanced measure that takes into account both precision and recall, making it a reliable metric for imbalanced classes. These metrics are crucial for researchers to understand the strengths and weaknesses of their classification models. Therefore, it is important to carefully evaluate these metrics and interpret the results to make informed decisions about model selection and improvement. Precision (ρ), Recall (τ) and F1-Score (F_1) can be defined in terms of true positive instance (ϵ), true negative instances (η), false positive instances (ζ), and false negative instance (θ) as follows.

$$\rho = \frac{\epsilon}{\epsilon + \zeta}; \quad \tau = \frac{\epsilon}{\epsilon + \theta}; \; F_1 = 2 \cdot \frac{\rho \cdot \tau}{\rho + \tau} \tag{39}$$

Table 3 tabulates the Precision, Recall and F1-Scores for all the algorithms. While PQSVC - C = 1000 has a balanced performance with the highest values across the 3 metrics:



(a) PegasosQSVC (C = 1000) (b) PegasosQSVC (C = 100) (c) PegasosQSVC (C = 10) (d) PegasosQSVC (C = 1)

(e) PegasosQSVC (C = 0.1) (f) PegasosQSVC (C = 0.01) (g) VQC - RealAmplitudes (h) VQC - EfficientSU2

FIGURE 4. Confusion matrix heatMap for algorithms under analysis.

TABLE 3. Precision, recall, and F1-score values for all QML algorithms.

Algorithm	Precision	Recall	F1-Score
VQC - RealAmplitudes	94.63	94.56	94.555
VQC - EfficientSU2	95.1	95.04	95.04
PQSVC - C = 1000	**97.96**	**97.96**	**97.94**
PQSVC - C = 100	87.71	85.94	86.81
PQSVC - C = 10	89.67	87.08	87.75
PQSVC - C = 1	95.57	95.53	95.52
PQSVC - C = 0.1	89.45	85.73	85.29
PQSVC - C = 0.01	88.74	87.91	87.81

Precision-Recall-F1Score $\sim$ 97.94%, VQC and it's ansatz variation and PQSVC with C $=$ 1000 and C $=$ 1 has all metrics almost equal with average difference Δ 0.065. Comparatively, PQSVC and its' C variations (C $=$ 100, C $=$ 10, C $=$ 0.1 and C $=$ 0.01) have a larger marginal difference of Δ_{100} 1.77, Δ_{10} 2.59, $\Delta_{0.1}$ 3.72 and $\Delta_{0.01}$ 0.83 for the precision recall values. Out of all algorithms, PQSVC C $=$ 0.1 has the lowest F1-Score, from which it is evident that the algorithm has a much lower performance with low recall, meaning weak ability to Distinguish Non-Hazardous Asteroids from PHAs. FIGURE 5c, 5b and 5d depict the Table 3 of Metrics graphically.

3) ACCURACY

The accuracy metric is an essential measure to evaluate the performance of a classification algorithm for identifying hazardous asteroids. It is calculated as the ratio of the correctly classified hazardous (ϵ), and non-hazardous (η) asteroids to the total number of asteroids under consideration ($\epsilon + \zeta + \eta + \theta$) [52], [53]. Accurate classification is critical since misclassifying a non-hazardous asteroid as hazardous can result in unnecessary disruptions while failing to recognize a hazardous asteroid can be disastrous. The mathematical expression for accuracy, denoted by the Greek symbol κ, is provided below.

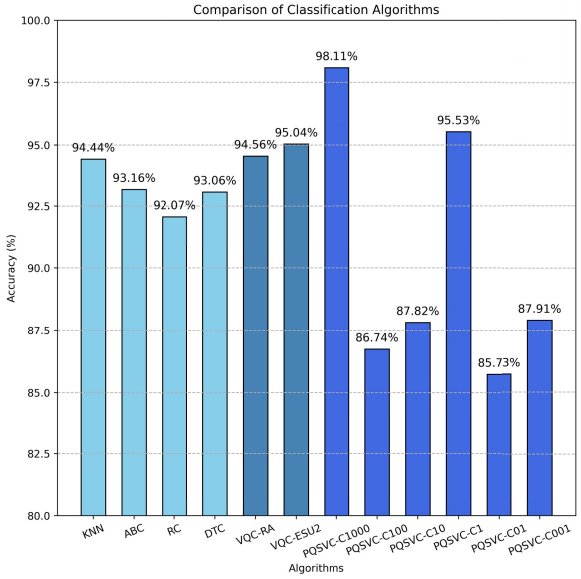
$$\kappa = \frac{\epsilon + \eta}{\epsilon + \zeta + \eta + \theta} \tag{40}$$

Equation 40 shows the ratio of the correctly classified hazardous and non-hazardous asteroids to the total number of asteroids. The symbol κ represents accuracy and provides insight into how effectively a model performs, which helps compare different models for predicting the state of an asteroid being hazardous.

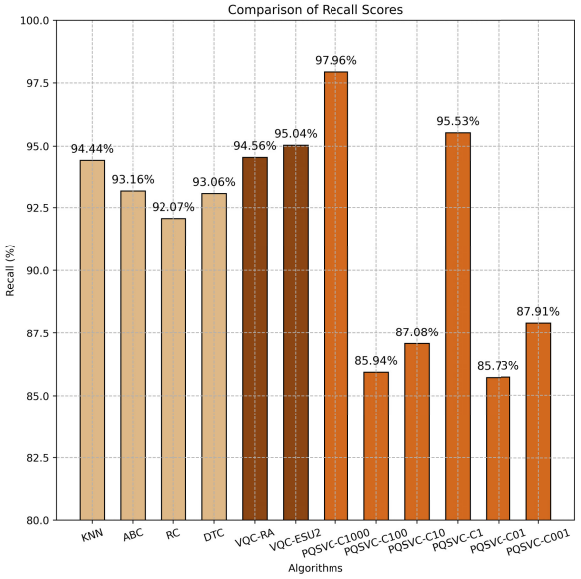
Results obtained for the eight different models indicate that the VQC algorithm performs very well with variations in the ansatz with accuracy values of 94.56% and 95.04% for RealAmplitudes ansatz and EfficientSU2 ansatz, respectively. For PegasosQSVC algorithm there is large variation $\triangle=$ 12.23 between the highest performing model (PegasosQSVC c $=$ 1000, Accuracy $=$ 98.11%) and other variations(PegasosQSVC c $=$ 0.01, 0.1, 1, 10, 100), with accuracy values 87.91%, 85.73%, 95.53%, 86.74%, and 87.82% respectively.

C. INTERPRETABILITY OF QML MODELS

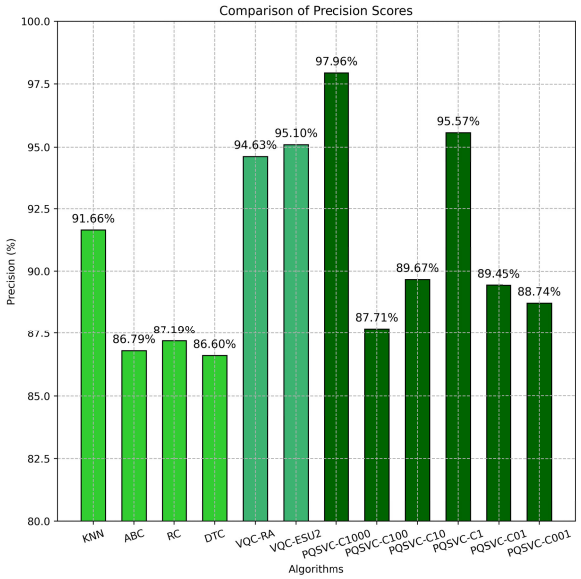
By maintaining an analogy between the VQC and PQSVC Models with classical counterparts, we can establish a working relationship between Neural Networks, Optimization techniques and Kernel tricks to explain how the Quantum algorithm can gain an advantage over the Classical ML algorithms. While QML models do not possess a direct metric for evaluating interpretability, we can delve into their underlying principles to understand how they make predictions. In the



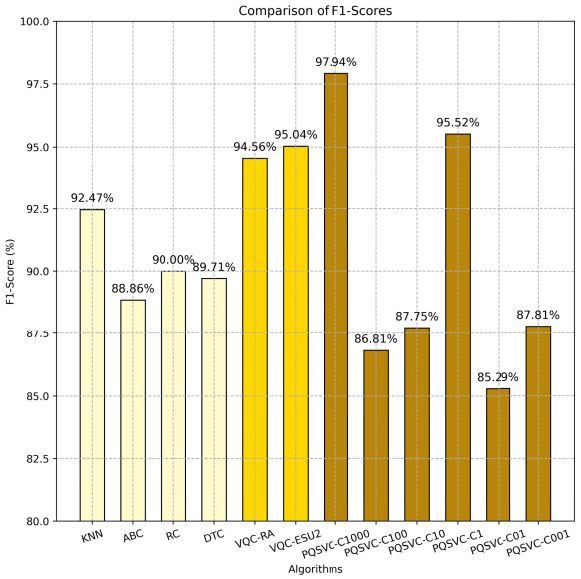
(a) Accuracy comparison between ML and QML models



(b) Recall comparison between ML and QML models



(c) Precision comparison between ML and QML models



(d) F1-score comparison between ML and QML models

FIGURE 5. Comparison between ML and QML models.

case of VQC, the quantum deep learning approach to classification involves parameter variations of the qubits within the quantum circuit. These variations can be interpreted as playing a role analogous to the activation values of neurons in classical deep neural networks. By applying predefined functions, the probabilities of quantum states within the circuit can be explicitly translated into the output values of the last layer. These output values, in turn, enable the prediction of classical labels or values. However, it's important to note that the exact choice of predefined functions can vary depending on the specific application or problem domain.

On the other hand, PQSVC shares similarities with its classical counterpart in terms of its kernel. In classical Support Vector Machines (SVM), the kernel function is responsible for transforming the input data into a high-dimensional feature space, where the linear separation between classes can be achieved. In PQSVC, the kernel function is based on the principles of quantum mechanics, harnessing the intrinsic high-dimensional properties offered by quantum systems. While the exact implementation and selection of quantum kernels may vary, the underlying idea remains consistent: leveraging quantum effects to efficiently map the input

data into a higher-dimensional space, potentially leading to improved classification performance. While interpretability in the traditional sense refers to the ability to understand and control a model's decision-making process, it is important to note that QML models, including VQC and PQSVC, are still nascent. The interpretability of QML models is currently limited due to the complex nature of quantum states and transformations. These models operate in a quantum realm that lacks direct classical analogs, making it challenging to provide explicit interpretations for their predictions.

D. DISCUSSION

The results comprise eight QML models, each with varying hyperparameters and performance metrics, compared with 4 ML models.

The first four algorithms are machine learning algorithms - Decision Tree, K-Nearest Neighbors (KNN), Adaboost Classifier, and Ridge Classifier - on the original dataset. The average accuracy achieved by these models was 93.25%, which mirrors the average recall score and an average F1-score of 90.26% and a precision score of 88.06%; a problem became apparent when these models were applied to a smaller subset of data initially prepared for Quantum Machine Learning (QML). The models showed a strong tendency to overfit, performing exceptionally well on the training data to the point of memorizing it, resulting in nearly 100% performances across all metrics.

Overfitting like this is a serious issue, as the models are less effective when presented with new, unseen data. In contrast, our initial tests with QML on the same data subset showed less evidence of overfitting, suggesting that QML models may perform better when generalizing to new data. These findings provide solid evidence for our decision to shift our focus to QML, supporting the argument that traditional machine learning algorithms are reaching their limits and newer, more advanced techniques are required.

The following two models are quantum-classical hybrid models, using the Variational Quantum Classifier (VQC) algorithm with different ansatzes and optimizers. The EfficientSU2 ansatz with COBYLA optimizer achieved an accuracy of 93.71%, while the RealAmplitudes ansatz with the same optimizer achieved a slightly better accuracy of 94.56%. Both Models have significantly fewer Type I (31 and 13) and Type II (14 and 39) errors and almost balanced sensitivity and specificity values, indicating an unbiased model. Both models' precision, recall, and f1-score values are also relatively high and balanced, indicating that they can effectively predict the hazardous asteroid without bias toward the non-hazardous floating object classification.

Lastly, the six models are quantum models using the classical Pegasos algorithm with the FidelityQuantumKernel and ZFeatureMap. These models vary in the value of the regularization parameter C. The model with $C = 1000$ achieved the best performance with an accuracy of 98.11% and high precision and recall values for both classes @0.9796 and @0.9794, respectively. The model with $C = 1$ had lower accuracy of 95.53% but maintained a balanced precision and recall for both classes @0.9557 and @0.9553. However, the models with $C = 0.1$ and $C = 0.01$ performed poorly, with accuracy values of 85.73% and 87.91%, respectively.

The Ansatz used in the quantum models, RealAmplitudes, and EfficientSU2, are different parameterized quantum circuits used to prepare the quantum state in the quantum-classical hybrid model. The choice of ansatz affects the expressivity of the circuit and how well it can fit the data. In this case, the EfficientSU2 ansatz, with its higher number of parameters, performed better than the RealAmplitudes ansatz.

The hyperparameter C controls the regularization strength in the PegasosQSVC algorithm. Higher values of C impose stricter boundaries on the decision boundary, which can lead to overfitting if the dataset is small or noisy. Lower values of C allow more misclassifications, which can result in underfitting if the dataset is complex. As the results show, the optimal value of C depends on the dataset's characteristics and should be chosen through experimentation. Comparing sensitivity over specificity, $c = 1000$ performed the highest out of other c-value variations.

The precision and recall values for each model provide insights into how well the model performs for each class. Precision measures the proportion of true positives among all positive predictions, while recall measures the proportion of true positives among all actual positive instances. High precision and recall values indicate that the model can effectively classify the data without being biased toward either class. However, a significant difference between precision and recall values for a specific class can indicate that the model struggles to identify that class, and improvements must be made.

In conclusion, the results indicate that the VQC algorithm can effectively classify data, and the choice of ansatz and optimizer can significantly affect the model's performance. The PegasosQSVC algorithm with the FidelityQuantumKernel and ZFeatureMap also achieved good results, with the optimal value of C depending on the dataset's characteristics. The precision and recall values provide valuable insights into the model's performance, especially in identifying the minority class, and should be considered when evaluating classification models.

VI. CONCLUSION AND FUTURE SCOPE

In conclusion, existing machine-learning approaches for predicting asteroid hazards are often resource-intensive, necessitating substantial computational time and effort. These methods, while effective, often grapple with efficiency, particularly in accurately identifying potential asteroid threats in a vast and rapidly expanding dataset. This limitation underscores the imperative for exploring more potent and efficient computational strategies such as Quantum Machine Learning. To address this issue, we proposed a Quantum Machine Learning-based methodology that leverages the quantum properties of data to improve the accuracy and

precision of asteroid classification. Our proposed method using Variational Quantum Circuits and PQSVC algorithms outperformed the existing techniques, achieving an accuracy of 98.11% and an average F1-score of 92.69%. Furthermore, incorporating image-based analysis in our future scope can enhance the accuracy and efficiency of the proposed work. Integrating our QML-based asteroid hazard prediction methodology into the actual system can aid in the real-time detection and mitigation of potential risks posed by asteroids, ultimately ensuring the safety of humans and biodiversity.

In future work, we will extend the proposed work to involve clustering of similar characteristic objects in 3d space, which defines to be complex with classical computing wherein the intrinsic nature of qubit to 3d-mapping based on qubit probability would largely help overcome and aid in the progress of the task, exciting to expanding on the advanced applications of QML and astronomy.

REFERENCES

[1] J. Aguirre et al., "Observing the evolution of the universe," 2009, *arXiv:0903.0902*.
[2] D. Jewitt, A. Moro-Martìn, and P. Lacerda, "The Kuiper belt and other debris disks," in *Astrophysics in the Next Decade: The James Webb Space Telescope and Concurrent Facilities*. Cham, Switzerland: Springer, 2009, pp. 53–100.
[3] Z. Landsman, "The physical properties and composition of main-belt asteroids from infrared spectroscopy," Electronic Theses and Dissertations, Univ. Central Florida, Gainesville, FL, USA, 2017.
[4] Q. Ye et al., "Toward efficient detection of small near-Earth asteroids using the Zwicky Transient Facility (ZTF)," *Publications Astronomical Soc. Pacific*, vol. 131, no. 1001, Jul. 2019, Art. no. 078002.
[5] C. F. Chyba, P. J. Thomas, and K. J. Zahnle, "The 1908 Tunguska explosion: Atmospheric disruption of a stony asteroid," *Nature*, vol. 361, no. 6407, pp. 40–44, Jan. 1993.
[6] A. W. Harris and P. W. Chodas, "The population of near-Earth asteroids revisited and updated," *Icarus*, vol. 365, Sep. 2021, Art. no. 114452.
[7] L. Denneau et al., "The Pan-STARRS moving object processing system," *Publications Astronomical Soc. Pacific*, vol. 125, no. 926, p. 357, 2013.
[8] M. Granvik, A. Morbidelli, R. Jedicke, B. Bolin, W. F. Bottke, E. Beshore, D. Vokrouhlický, D. Nesvorný, and P. Michel, "Debiased orbit and absolute-magnitude distributions for near-Earth objects," *Icarus*, vol. 312, pp. 181–207, Sep. 2018.
[9] M. Granvik, A. Morbidelli, R. Jedicke, B. Bolin, W. F. Bottke, E. Beshore, D. Vokrouhlický, M. Delbò, and P. Michel, "Super-catastrophic disruption of asteroids at small perihelion distances," *Nature*, vol. 530, no. 7590, pp. 303–306, Feb. 2016.
[10] M. Delbo, M. Mueller, J. P. Emery, B. Rozitis, and M. T. Capria, "Asteroid thermophysical modeling," *Asteroids IV*, vol. 1, pp. 107–128, Aug. 2015.
[11] M. Colazo, A. Alvarez-Candal, and R. Duffard, "Zero-phase angle asteroid taxonomy classification using unsupervised machine learning algorithms," *Astron. Astrophys.*, vol. 666, p. A77, Oct. 2022.
[12] M. Bhavin, S. Tanwar, N. Sharma, S. Tyagi, and N. Kumar, "Blockchain and quantum blind signature-based hybrid scheme for healthcare 5.0 applications," *J. Inf. Secur. Appl.*, vol. 56, Feb. 2021, Art. no. 102673.
[13] V. K. Ralegankar, J. Bagul, B. Thakkar, R. Gupta, S. Tanwar, G. Sharma, and I. E. Davidson, "Quantum cryptography-as-a-service for secure UAV communication: Applications, challenges, and case study," *IEEE Access*, vol. 10, pp. 1475–1492, 2022.
[14] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, pp. 195–202, Sep. 2017.
[15] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019.
[16] S. Shalev-Shwartz, Y. Singer, and N. Srebro, "Pegasos: Primal estimated sub-gradient solver for SVM," in *Proc. 24th Int. Conf. Mach. Learn.*, Jun. 2007, pp. 807–814.
[17] N. Petrov, L. Sokolov, E. Polyakhova, and K. Oskina, "Predictions of asteroid hazard to the Earth for the 21st century," *AIP Conf. Proc.*, vol. 1959, no. 1, May 2018, Art. no. 040012.
[18] V. Carruba, S. Aljbaae, and A. Lucchini, "Machine-learning identification of asteroid groups," *Monthly Notices Roy. Astronomical Soc.*, vol. 488, no. 1, pp. 1377–1386, Sep. 2019.
[19] K. J. McIntyre, "Applying machine learning to asteroid classification utilizing spectroscopically derived spectrophotometry," Dept. Space Stud., Univ. North Dakota, Grand Forks, ND, USA, 2019.
[20] V. Pasko, "Prediction of orbital parameters for undiscovered potentially hazardous asteroids using machine learning," in *Stardust Final Conference* (Asteroids and Space Debris Engineering and Science). Cham, Switzerland: Springer 2018, pp. 45–65.
[21] M. Chhibber, M. Bhatia, A. Chaudhary, and C. Stewart, "Comparing the efficacy of machine learning models on potentially hazardous objects," in *Proc. 2nd Int. Conf. Technol. Advancements Comput. Sci. (ICTACS)*, Oct. 2022, pp. 725–730.
[22] R. N. Ranaweera and T. Fernando, "Prediction of potentially hazardous asteroids using deep learning," in *Proc. 2nd Int. Conf. Adv. Res. Comput. (ICARC)*, Feb. 2022, pp. 31–36.
[23] O. Simeone, "An introduction to quantum machine learning for engineers," *Found. Trends Signal Process.*, vol. 16, nos. 1–2, pp. 1–223, 2022.
[24] W. Joon Yun, H. Baek, and J. Kim, "Projection valued measure-based quantum machine learning for multi-class classification," 2022, *arXiv:2210.16731*.
[25] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Phys. Rev. Lett.*, vol. 122, no. 4, Feb. 2019, Art. no. 040504.
[26] H. Park and Y.-C. Jeong, "Quantum algorithm for the two-body problem in quantum mechanics," 2020, *arXiv:2011.01319*.
[27] A. Leverrier and M. Weigand, "Quantum algorithms for space situational awareness," 2021, *arXiv:2109.05728*.
[28] E. Zahedinejad, M. Schuld, and N. Killoran, "Quantum variational algorithms for the Kepler problem," 2020, *arXiv:2012.09242*.
[29] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 326–327, Sep. 1995.
[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
[31] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, May 2003.
[32] D. Castelvecchi, "Can we open the black box of AI?" *Nature*, vol. 538, no. 7623, pp. 20–23, Oct. 2016.
[33] S. Barocas and A. D. Selbst, "Big data's disparate impact," *California Law Review*, vol. 104, pp. 671–732, Jun. 2016.
[34] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*. Cambridge, MA, USA: MIT Press, 2008.
[35] S. Jerbi, L. J. Fiderer, H. Poulsen Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, "Quantum machine learning beyond kernel methods," *Nature Commun.*, vol. 14, no. 1, p. 517, Jan. 2023.
[36] M. K. Gupta, M. Romaszewski, and P. Gawron, "Potential of quantum machine learning for processing multispectral Earth observation data," *TechRxiv*, pp. 1–7, 2023.
[37] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, "Generalization in quantum machine learning from few training data," *Nature Commun.*, vol. 13, no. 1, p. 4919, Aug. 2022.
[38] M. Schuld and N. Killoran, "Is quantum advantage the right goal for quantum machine learning?" *PRX Quantum*, vol. 3, no. 3, Jul. 2022, Art. no. 030101.
[39] R. Dilip, Y.-J. Liu, A. Smith, and F. Pollmann, "Data compression for quantum machine learning," *Phys. Rev. Res.*, vol. 4, no. 4, Oct. 2022, Art. no. 043007.
[40] J. Kubečka, A. S. Christensen, F. R. Rasmussen, and J. Elm, "Quantum machine learning approach for studying atmospheric cluster formation," *Environ. Sci. Technol. Lett.*, vol. 9, no. 3, pp. 239–244, Mar. 2022.
[41] W. Li and D.-L. Deng, "Recent advances for quantum classifiers," *Sci. China Phys., Mech. Astron.*, vol. 65, no. 2, Feb. 2022, Art. no. 220301.
[42] J. D. Martín-Guerrero and L. Lamata, "Quantum machine learning: A tutorial," *Neurocomputing*, vol. 470, pp. 457–461, Jan. 2022.
[43] J. Roa, D. Farnocchia, and S. R. Chesley, "A novel approach to asteroid impact monitoring," *Astronomical J.*, vol. 162, no. 6, p. 277, Dec. 2021.

[44] M. S. Hossain and M. A. Zabed, "Machine learning approaches for classification and diameter prediction of asteroids," in *Proc. Int. Conf. Inf. Commun. Technol. Develop.*, M. Ahmad, M. S. Uddin, and Y. M. Jang, Eds. Singapore: Springer, 2023, pp. 43–55.
[45] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.
[46] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[47] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *J. Jpn. Soc. Artif. Intell.*, vol. 14, nos. 771–780, p. 1612, Sep. 1999.
[48] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[49] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.
[50] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, pp. 413–422.
[51] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, Dec. 1997.
[52] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *Advances in Artificial Intelligence*. Cham, Switzerland: Springer, 2006, pp. 1015–1021.
[53] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 168–192, Jan. 2021.

RUSHIR BHAVSAR is currently pursuing the B.Tech. degree in computer science and engineering with the Institute of Technology, Nirma University, Ahmedabad, India. His research interests include quantum computing, federated learning, machine learning, and deep learning.

NILESH KUMAR JADAV (Student Member, IEEE) received the bachelor's and M.Tech. degrees from Gujarat Technological University (GTU), Gujarat, India, in 2014 and 2018, respectively. He is a full-time Ph.D. Research Scholar with the Department of Computer Science and Engineering, Nirma University, Ahmedabad, Gujarat. He has authored/coauthored publications (including papers in SCI-indexed journals and IEEE ComSoc-sponsored international conferences). Some of his research findings are published in top-cited journals and conferences, such as IEEE Transactions on Industrial Informatics, IEEE Transactions on Network and Service Management, IEEE INFOCOM, IEEE ICC, and IJCS. His research interests include artificial intelligence, network security, 5G communication networks, and blockchain technology. He is an Active Member of the ST Research Laboratory (www.sudeeptanwar.in).

UMESH BODKHE received the B.E. degree in computer engineering from the University of Pune and the M.Tech. degree in computer science and engineering from Jawaharlal Nehru Technological University Hyderabad. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Nirma University, Ahmedabad, Gujarat, India. He is an Assistant Professor with the Department of Computer Science and Engineering, Institute of Technology, Nirma University. He has more than eight years of teaching experience. His research areas are security in IoT, blockchain, IoV, and 5G. He has taught various courses, such as communication engineering, blockchain technology, data privacy, digital systems, data communications, and operating systems at the undergraduate and postgraduate levels. He is a Life Member of ISTE. He has authored or coauthored more than 30 research papers in leading SCI journals and top core IEEE COMSOC A* conferences. Some of his top-notch findings are published in reputed SCI journals, such as *Transactions of Emerging Telecommunication Technologies* (John Wiley), *Transactions of Network Science and Engineering*, *Software - Practice and Experience* journal (John Wiley), *IET Software* journal, IEEE Access, MDPI, IEEE Globcomm, IEEE Infocomm, IEEE International Conference on Computer, Information and Telecommunication Systems (CITS-2019, 2020), Proceedings of Springer First International Conference on Computing, Communications, and Cyber-Security (IC4S-2019 and IC4S2020). He is also a Reviewer of reputed SCI/Scopus journals, such as IEEE Transactions on Neural Networks and Learning Systems, IEEE Sensors Journal, *Transactions on Emerging Telecommunications Technologies*, *Journal of Information & Knowledge Management*, *Signal, Image and Video Processing* journal, *MethodsX* journal, *Peer-to-Peer Networking and Applications*, *Telematics and Informatics*, *Sustainability* (MDPI), IEEE Access, and *International Journal of Information Technology & Decision Making*.

RAJESH GUPTA (Member, IEEE) received the B.E. degree from the University of Jammu, Jammu, India, in 2008, the master's degree in technology from Shri Mata Vaishno Devi University, Jammu, in 2013, and the Ph.D. degree in computer science and engineering from Nirma University, Ahmedabad, Gujarat, India, under the supervision of Dr. Sudeep Tanwar. He is an Assistant Professor with Nirma University. He has authored/coauthored some publications (including papers in SCI-indexed journals and IEEE ComSoc-sponsored international conferences). Some of his research findings are published in top-cited journals and conferences, such as IEEE Transactions on Industrial Informatics, IEEE Transactions on Network and Service Management, IEEE Transactions on Network Science and Engineering, IEEE Transactions on Green Communications and Networking, IEEE Transactions on Computational Social Systems, *IEEE Network Magazine*, IEEE Internet of Things Journal, *IEEE Internet of Things Magazine*, *Computer Communications*, *Computer and Electrical Engineering*, *International Journal of Communication Systems* (Wiley), *Transactions on Emerging Telecommunications Technologies* (Wiley), *Physical Communication* (Elsevier), IEEE ICC, IEEE INFOCOM, IEEE GLOBECOM, and IEEE CITS. His H-index is 29 and i10-index is 47. His research interest includes device-to-device communication, network security, blockchain technology, 5G communication networks, and machine learning. He was a recipient of the Doctoral Scholarship from the MeitY, Government of India, under the Visvesvaraya Ph.D. Scheme, and the Student Travel Grant from WICE-IEEE to attend IEEE ICC 2021 held in Canada. He was awarded the Best Research Paper Award from IEEE SCIoT 2022, IEEE ECAI 2021, IEEE ICCCA 2021, and IEEE IWCMC 2021. His name has been included in the list of Top 2% scientists worldwide published by Stanford University, USA, in 2021 and 2022. He was felicitated by Nirma University for their research achievements, from 2019 to 2020 and from 2021 to 2022. He is also an Active Member of the ST Research Laboratory.

SUDEEP TANWAR (Senior Member, IEEE) received the B.Tech. degree from Kurukshetra University, India, in 2002, the M.Tech. degree (Hons.) from Guru Gobind Singh Indraprastha University, Delhi, India, in 2009, and the Ph.D. degree with a specialization in wireless sensor networks, in 2016. He is currently working as a Professor with the Department of Computer Science and Engineering, Institute of Technology, Nirma University, India. He is also a Visiting Professor with Jan Wyzykowski University, Polkowice, Poland; and the University of Pitesti, Pitesti, Romania. He is also leading the ST Research Laboratory, where group members are working on the latest cutting-edge technologies. He has authored two books, edited 13 books, and more than 250 technical articles, including top journals and top conferences, such as IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE WIRELESS COMMUNICATIONS, *IEEE Network Magazine*, ICC, GLOBECOM, and INFOCOM. He initiated the research field of blockchain technology adoption in various verticals, in 2017. His H-index is 63. He actively serves his research communities in various roles. His research interests include blockchain technology, wireless sensor networks, fog computing, smart grids, and the IoT. He is a Final Voting Member of the IEEE ComSoc Tactile Internet Committee, in 2020. He is also a member of CSI, IAENG, ISTE, CSTA, and the Technical Committee on Tactile Internet of the IEEE Communication Society. He has been awarded the Best Research Paper Award from IEEE IWCMC-2021, IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has served many international conferences as a member of the organizing committee, such as the Publication Chair for FTNCT-2020, ICCIC 2020, and WiMob2019; a member for the Advisory Board for ICACCT-2021 and ICACI 2020; the Workshop Co-Chair for CIS 2021; and the General Chair for IC4S 2019 and 2020 and ICCSDF 2020. He is serving on the Editorial Board for *Computer Communications*, *International Journal of Communication Systems*, and *Security and Privacy*.

GULSHAN SHARMA received the B.Tech., M.Tech., and Ph.D. degrees. He is currently a Senior Lecturer with the Department of Electrical Engineering Technology, University of Johannesburg. He is also a Y-rated Researcher with NRF South Africa. His research interests include power system operation and control and the application of AI techniques to power systems. He is an Academic Editor of *International Transactions on Electrical Energy System* (Wiley) and a Regional Editor of *Recent Advances in Electrical & Electronic Engineering* (Bentham Science).

PITSHOU N. BOKORO (Member, IEEE) received the M.Phil. degree in electrical engineering from the University of Johannesburg, Johannesburg, South Africa, in 2011, and the Ph.D. degree in electrical engineering from the University of the Witwatersrand, in 2016. He is currently an Associate Professor with the University of Johannesburg. His research interests include modeling and reliability prediction of insulating materials and dielectrics, power quality, and renewable energies. He is a Senior Member of the South African Institute of Electrical Engineers.

RAVI SHARMA is a Professor with the Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun, India. He is passionate in the field of business analytics and worked in various MNCs as a Leader of various software development groups. He has contributed various articles in the area of business analytics, prototype building for startups, and artificial intelligence. He is leading academic institutions as a Consultant to uplift research activities in inter-disciplinary domains.