IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

Enhancing Privacy and Efficiency in Ride-Matching Systems through Federated Learning

Wendy Nicola Tomusoni, Tashinga Bwanali, Wagner Marques, Aditya Dayal Tyagi
Department of Computer Science and Engineering
School of Engineering and Technology, Sharda University, Greater Noida, India
2022808783.wendy@ug.sharda.ac.in, 2022809445.tashinga@ug.sharda.ac.in, 2022811805.wagner@ug.sharda.ac.in
adityadayaltyagi@gmail.com

Abstract: *Ride-matching companies gather user data in real-time about their locations and preferred travel routes, as well as their previously taken rides. However, this type of data centralization presents significant privacy concerns. Cyber criminals can target such centralized databases, which would then expose users to data leaks as well as misuse of confidential information. A possible solution to the aforementioned problem is provided by federated learning, which decentralizes the learning process by keeping user data on their devices and only sending updates to the model. While making it more difficult for unauthorized individuals to access personal data, federated learning still provides high quality services for ride matching. Our work demonstrates that privacy preserving methods like secure aggregation and differential privacy provide data protection alongside efficient performance. Secure aggregation collects updates from multiple users prior to transmission to hide the identity of individual users, while differential privacy obfuscates the identity of the user by adding noise, making them much harder to extract. Both methods allow for the construction of federated learning systems that are privacy aware and do not use traditional methods of machine learning. Though, certain challenges still exist such as disparity in data, communication costs, and the variety in devices. While federated learning makes use of out-of-the-box computer systems, the varying capacity of the hardware and the quality of the networks may affect the learning process. To solve these problems, the model and communication design optimization has to be done for effective operation. Even with these restrictions set, in comparison to other methods, federated learning is superior considering the fact that it protects privacy greatly, improves the intelligence of the system, and provides a friendly user experience. Doing so reduces the chances of privacy infringement which builds trust in ride-sharing apps, making them easier to use and safer*

Keywords: Federated Learning, Ride-Matching, Privacy, Decentralized AI, Secure Data Processing

I. INTRODUCTION

The introduction of ridesharing companies such as Uber, Bolt, and Lyft has made transportation effortless and achievable by using technology for real-time optimization of passenger- driver routing and matching along with the affordable pricing of services. On the flip side, Fizz, their excessive collection and storages of users' data like payment details, trip history, location and many more provide major concerns in regards to privacy. Centralized systems, when all information of users is stored at one place, is very prone to attacks which was evident with the Uber 2016 hack which leaked a staggering 57 million user's data [3]. This has increased the need for more advanced systems such as Federated Learning (FL), which is able to decenteralize data processing by keeping sensitive data on users devices, and only sending model updates to a central server in a encrypted format [4].

By decentralizing the collection of user data, FL reduces the chances of massive data breaches occurring. Traditional models that depend on data being stored on a central server in a secured location suffer from security issues. Rather than storing sensitive data, FL runs computations for model parameters on the users' devices [1]. The data also remains confidential, which results in users having greater confidence in ride-sharing applications and encourages them to share their data even further [6]. Although rewarding, FL is also challenged by the issues of device variety, data imbalance,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

and high cost of communication. Diversity in devices can decrease the effectiveness of model training, while skewed data distribution may result in inaccurate [9]. Moreover, increasing latency on the network as a result of constant model updates between the server and devices can slow down the quality of service [15].

This study looks at the use of Federated Learning (FL) on data privacy enhancement for ride-matching platforms and assesses its effectiveness against centralized systems. It analyzes the prospects of FL improving security, efficiency, and scalability, while overcoming the obstacles of implementation such as device heterogeneity and high communication costs. The study also focuses on the use of more sophisticated technologies such as blockchain and data protection homomorphic encryption [18, 19]. Alongside these issues, legal and ethical aspects such as GDPR and CCPA compliance regulations are analyzed with special attention to the fairness and openness FL frameworks [16, 20].

This study investigates the practicality of device-level Federated Learning in large-scale ride-sharing systems and aims to address the intersection of privacy, security, and performance. The results of the study will respond to the emerging need for privacy-preserving AI in urban mobility while guaranteeing user data protection and quality of service.

II. LITERATURE SURVEY

Just like balancing service efficiency and privacy, user data protection, even in McGregor and colleagues work, still comes with privacy concerns that are directly proportional to the level of service provided. Ride matching services make it easy for drivers and riders to communicate and interact instantaneously, while users data such as location, payment methods and even travel history is greatly safeguarded. In centralized databases, these sensitive user information comes under great protection, however, these dry systems are prone to breaches jeopardizing user security and paving the way to identity theft and espionage activities. This causes users to be skeptical in sharing their sensitive data which puts companies in a tight spot as they have to find a middle ground.

The introduction of encryption techniques helps in alleviating individual concerns such as enabling users to mask or shield sensitive parts of their data. It also brings forth statistical noise in identifiable datasets in a manner that protects individual identities while still retaining usefulness, referred to as differential privacy. Nevertheless, these methods have drawbacks, such as decreased system speeds owing to claimed encryption or potential loss of accuracy in models that use identified privacy [6][7][13]. These limitations show the requirement for better solutions that meet privacy requirements without compromising efficiency, translating data into more practical insights.

The centralization of machine learning models aggravates privacy concerns because all data is kept in one place, increasing the potential for massive breaches [8]. Moreover, the growth of ride-matching services is accompanied by a massive amount of data that can strain system resources and result in delays and higher operational expenses [9]. Such difficulties have raised the demand for more secure and effective solutions, including Federated Learning (FL) [14].

FL differs from other models as it processes the information directly on the user's device and only sends an encrypted model update to the central server. Since no sensitive information leaves the device, privacy concerns are mitigated. FL is becoming popular in sensitive privacy domains such as health care and finance where protection of data is highly guarded [14]. In ride-matching services, location and travel data are stored on the user's device, instead of on the central server, thus enhancing privacy [16]. Additionally, it allows for more accurate predictions and an improved user experience as services can be tailored to an individual's preferences [16].

Nevertheless, there remain obstacles with FL such as device diversity, bias data, and communication overhead. Device heterogeneity can slow down model training because not all devices are equal in their processing power [15]. Imbalanced models are created because of data bias resulting from disproportionate data allocation, for example, urban users produce much more data than rural users [9]. Overhead communication costs also grow with the number of users, as frequent interaction of devices with the server leads to network inefficiency and delay [4].

With those constraints in mind, FL has successfully collaborated in healthcare and finance, assisting in secure collaboration and fraud detection without revealing sensitive information [3, 4]. Such accomplishments highlight FL's applicability in ride-matching services where privacy and security are critical. Although, there are challenges to tackle such as device fragmentation, data imbalance, and high communication latency before FL can be widely applied [19].

Considerations that are legal and ethical also matter. Compliance with laws like GDPR adds an additional layer of protection for users' data and privacy [20]. FL requires transparency, informed consent, and a way for users to request

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, April 2025 Impact Factor: 7.67

the revocation of their consent to share data. Addressing these issues is important in gaining trust and ensuring equity in FL-enabled systems.

FL holds great potential for solving problems related to privacy and security of ride-matching services. With FL's decentralized data processing, the risk of breaches is reduced while still enabling effective and efficient service delivery. Noteworthy, however, is the implementation of technical, legal, and ethical solutions that stand as obstacles. There should be more exploration and creativity applied to improve the use of FL in ride-matching systems so that people can enjoy better privacy, security, and usability.

III. EXISTING SOLUTIONS

3.1. AI-Based Anomaly Detection Systems

AI technology improves security in ride-sharing services by preventing fraud in real time through the detection of fake rides and unauthorized logins. These systems, however, need to be updated regularly to keep up with new developments in cyber threats. Ride-sharing services tend to prefer centralized data storage, implementing encryption via SSL or TLS to shield interactions between clients and servers, thus making it difficult for sensitive data to be intercepted while being transmitted. With these steps put in place, centralized systems still remain and are vulnerable to breaches. One attack may result in the compromise of millions of users accounts and highlight the need for more secure decentralized systems to be put in place.

3.2. Differential Privacy

Differential Privacy (DP) in context protects user data by allowing the addition of artificial noise which enables companies to analyze trends without the risk of exposing individual information (Dwork& Roth, 2014). In the context of ride-hailing, DP prevents the tracking of users by obscuring their location data. An example would be Uber which uses DP in the form of customers' routes being anonymized for route optimization (Erlingsson et al., 2019). Striking a balance between privacy and data quality remains a challenge on the DP side as added noise can often diminish the usefulness of the data (Geyer et al., 2017).

IV. PROPOSED SOLUTION: FEDERATED LEARNING

4.1. Federated learning approach

In this section, we suggest implementing Federated Learning (FL), a model whereby users' data is stored on devices while updates are forwarded to a central server (McMahan et al., 2017), to improve privacy and effectiveness in ride-matching systems. Because sensitive information, such as trip details and real-time positions, are not sent for direct all-in-one storage, FL reduces the risk of privacy breach from being exposed. With Federated Learning, local models can be trained on devices such as smartphones or in-car systems so long as only gradients (model parameters) are sent to the server. This is contrary to the standard FL approach which sends all data to a single location for processing (Yang et al., 2019). With this solution, user data will be protected while improving ride prediction accuracy at the same time.

4.2. System Architecture

The Federated Learning-based ride-matching system consists of three main components:

- **Global Model**: the central server holds the combined knowledge from all devices, which is called the global model. This model is continuously improved with new processed data from ride matching
- **Central Server**: The central server collects and aggregates model updates from all client devices. This aggregated model is refined to enhance the global ride-matching system, which is then redistributed to user devices for further local training.
- **Global Model**: Stored on the central server, the global model represents the combined knowledge gained from all participating devices. Periodic updates ensure that the model continues evolving as new ride-matching data is processed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

4.3. Data Flow and Privacy Mechanisms

The Federated Learning model processes ride-matching systems with a number of important attributes that are listed here:

User's Ride Preferences: Desired routes, types of vehicles, and pickup/drop-off locations of the user.

Historical Ride Data: Documents pertaining to past rides along with the timestamps, its frequency, and the time it took to complete them.

User Location: Data regarding user's location, which is anonymized with codename for privacy purposes, and is enabled for accurate ride-matching.

Federated Learning ensures that raw user data stays on the device, and only model updates are sent to the central server for merging. This approach maintains user privacy while enabling the global model to enhance its predictive accuracy.

To strengthen privacy, additional techniques need to be implemented:

Differential Privacy conceals sensitive data from model updates by adding irrelevant information that renders any individual data point useless.

Secure Aggregation ensures confidentiality of individual model updates, preventing the central server from distinguishing user data.

4.4. Model Operation and Workflow

The steps depicting the iterative cycle in the federated learning process is as follows:

1. **Initialization:** A global model is set as a baseline and sent out to all user devices.
2. **Local Training:** Each device gets to actively train a device specific model using their relevant data like trip data, location history, and ride preferences.
3. **Model Update**: After training, devices compute model gradients (updates to parameters) and securely transmit them to the central server.
4. **Aggregation**: The central server integrates updates from multiple devices to refine the global model.
5. **Global Model Update**: The improved model is redistributed to user devices, where the cycle repeats.

The model is able to self-improve accuracy with every iteration yielding better rides-matching predictions making the experience better and more suited to user's needs.

4.5. Evaluation Metrics

These metrics are put in place to assess the effectiveness of the federated learning based ride-matching system:

- **Matching Accuracy**: Determines how well the system does in matching users with appropriate rides.
- **Privacy Preservation**: Measures how well the system protects user information and in implementing privacy techniques such as Differential Privacy.
- **Efficiency**: Assesses the system's capability to provide real-time ride-matching with minimal computational and network overhead.
- Scalability: Estimates the number of users and devices the system can sustain without performance deterioration.

V. METHODOLOGY

This section explains in detail the FL methodology of ride-matching systems. In order to aid understanding, we provide visual diagrams depicting the system architecture, model training, and data flow. This part is organized in a way that FL is preferred over the centralized methods so that data processing is automated and enhances user privacy.

5.1 System Model & Architecture

This part describes how Federated Learning is implemented in ride-matching systems in detail. The system architecture has two key parts: the clients' devices (smartphones) and an intermediary server. The process starts with a client device training a local model with its confidential data. After training the local model, only the model updates which comprise of some weights and gradients are transmitted to the central server. With the help of Federated Averaging (FedAvg), the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

central server integrates those updates to enhance the global model. Subsequently, the central server sends out the improved global model to every client where it is stored and used to increase the accuracy of predictive responses to subsequent ride solicitations.

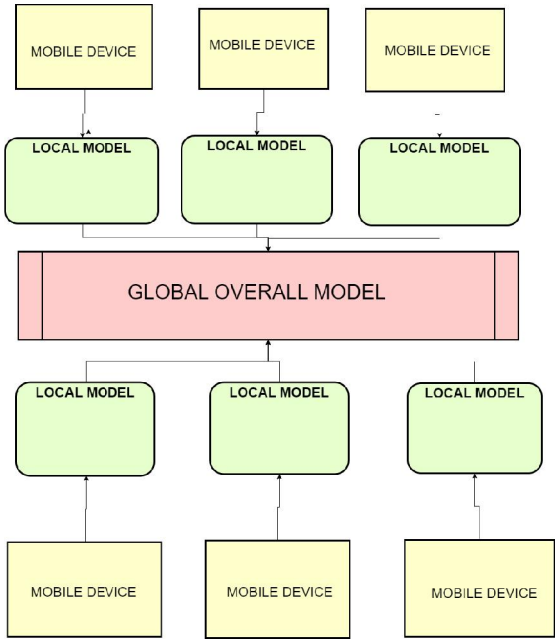

Fig.1: System Architecture

This diagram serves to illustrate the relationship of the user devices with the central server. It indicates the process in which several client devices (Mobile Devices 1-6) train local models on their private data and send encrypted model update information to the central server (Hosted Cloud Model). The central server improves the global model with the aggregated updates and redistributes it back to the clients. The diagram assists readers in fathoming the FL process's decentralized aspect and the model update process in detail.

Diagram Explanation:

- **Client Devices (Mobile Devices 1-6):** Each device trains a local model with their private ride data (e.g., trip history and settings). The local model initializes computations to determine weights and gradients for the accumulated data.
- **Central Server (Hosted Cloud Model):** The server collects all the encrypted model updates sent from the client devices and aggregates them with Federated Averaging (FedAvg). Then, the global model is updated and returned back to the clients.
- **Flow of Model Updates:** The diagram here demonstrates how model updates flow from client devices, to the central server, and back to the client devices without sharing any raw data.

5.2 Algorithm and Model Training

This segment describes the machine learning algorithms and models that have been applied in the study, focusing on their contribution towards optimizing ride-matching forecasts and user privacy through Federated Learning.

Neural Network Architecture

The individual model stored on each users' gadget is a neural network coded in PyTorch that comprises the following layers:

- **Input Layer**: Receives matched rides data which is semi-structured, for instance, the user's location, preferred trips, and previous trips.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

- **Hidden Layer**: Applies ReLU activation to detect ride-demand features.
- **Output Layer**: Provides ride-matching predictions.

The model performs training on the device using a stochastic gradient descent (SGD) algorithm that adjusts the predictions based on the interactions the user has with the device, and the data never leaves the device.

Federated Learning Process

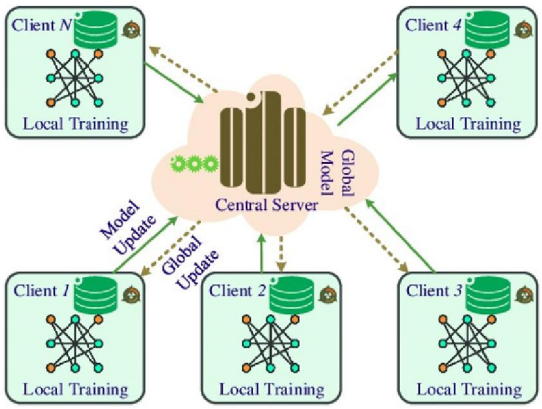

Fig. 2: Flow of Activities

Fig 2 shows the steps and their sequence in a Federated Learning (FL) system that functions in a loop:

1. **Local Training**: Each client device (Client 1, Client 2, Client 3, Client 4, Client N) trains a local model using their data without disclosing their raw data.
2. **Uploading the model**: After training concludes, each client uploads only the model updates, such as the weights and gradients, instead of the raw data to the central server.
3. **Model Improvement**: The central server receives the updates from all clients and improves the global model using these updates.
4. **Providing the model**: The global model that has been manipulated from the client updates is then sent back to all clients so they may train with further instructions.
5. **Refinement Looping**: This loop is executed several times until the global model reaches an optimum level of performance.

5.3. Local Model

Training is structured into these phases:

1. **Local Training**: In this phase, every device independently adjusts the model based on the ride records each individual possesses.
2. **Gradient Computation & Transmission**: The central server only receives model parameter changes and not any data.
3. **Secure Aggregation**: The server executes FedAvg, which helps in combining model changes from many devices.
4. **Global Model Refinement**: The devices get the updated global model, which increases the accuracy of the predictions further.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

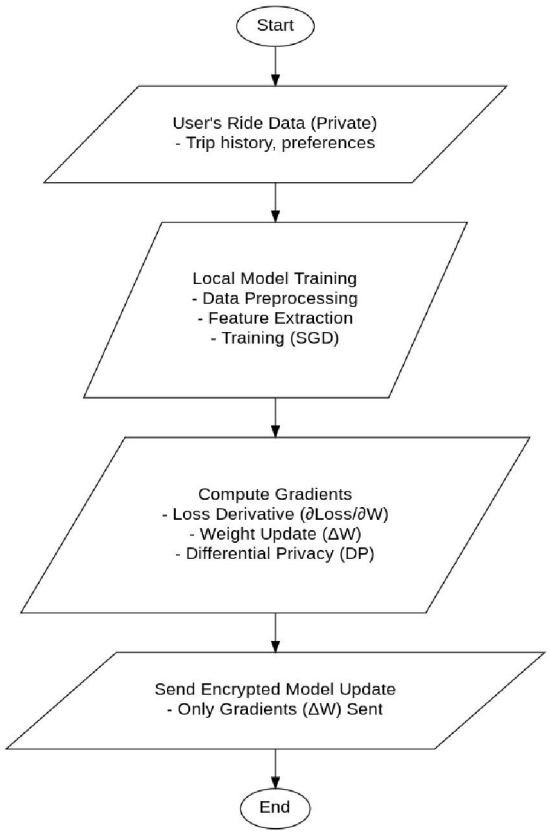

Fig.3: Local Model

This diagram illustrates the methodology of how ride information is processed on a user device for local model training privately. It captures the important constituents preceding the secure transfer of information to the central server.

Key Steps:

1. **Processing Private Ride Data:** The user's ride history and specific preferences remain stored on the device.
2. **Training the Local Model:** The model 'sees' the ride data, learns, and makes changes to its internal parameters.
3. The loss function, ∂Loss/∂W, is derived and is desecrated with noise for the sake of privacy using Differential Privacy(DP).
4. **Secure Transmission of Model Updates:** Encrypted model updates are transmitted to the central server without any identification information. Hence, privacy is maintained.

It helps in improving accuracy of the model while ensuring the data of the users is preserved.

5.4. Global Model

This illustration shows how the central server processes encrypted model updates received from client devices and improves the global model. As privacy is maintained, the system is able to constantly improve the predictions of the ride-matching system.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

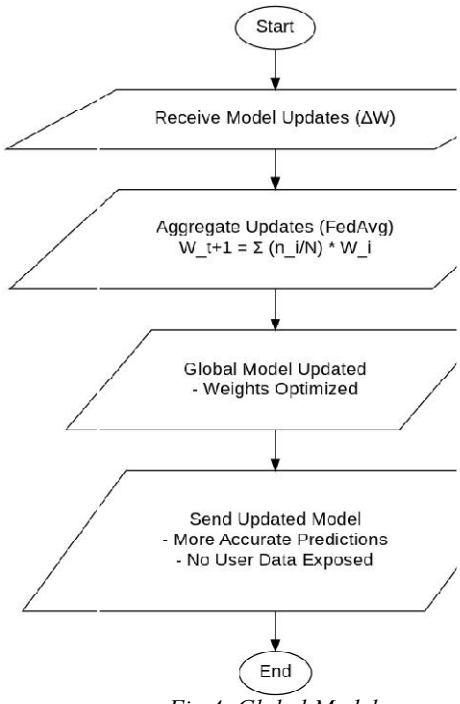

Fig 4: Global Model

Key Steps:

1. **Receiving Model Updates (ΔW):** The central server receives updates which are encrypted and sent from several client units.
2. **Aggregating Updates (FedAvg Algorithm):** The server executes the federated averaging algorithm where the updates are mixed and the global model is improved.
3. **Updating and Distributing the Global Model:** W_{t+1} is sent back to client devices which utilize the global model, enhancing ride-matching predictions with no user data exposure.

This approach is much more efficient in terms of system performance and user data privacy keeping the system secure.

5.5 Model Update

This diagram shows the complete flow of the model update in the Federated Learning system. It shows how the local model processes trip details, assesses historical ride data, and fits user's preferences. Instead of providing raw data, the only thing that gets sent to the central server are model updates in the form of gradients or weights, with additional optional encryptions or differential privacy to enhance security. These updates are aggregated by the central server in order to improve the ride matching AI. After the global model is refined, it is sent back to the users with advanced privacy features enabling better ride-matching prediction.

Diagram Explanation:

- **Local Model Processes:** The trip information is studied and augmented with other rides to devise the best matching options, and user-specific changes are incorporated to enhance the personalization capabilities.
- **Secure Model Updates:** Only modifications corresponding to update weights and the model's gradients are sent to the central server. These updates are optionally encrypted or added noise for privacy purposes.
- **Central Server Aggregates:** Aggregates and processes the received updates with Federated Averaging to improve the combined global ride-matching model in the server.
- **Updated Model Sent to Users:** The retrieved model is sent to the users and confidentiality of the user's data is still upheld, prompting more accurate ride matches.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

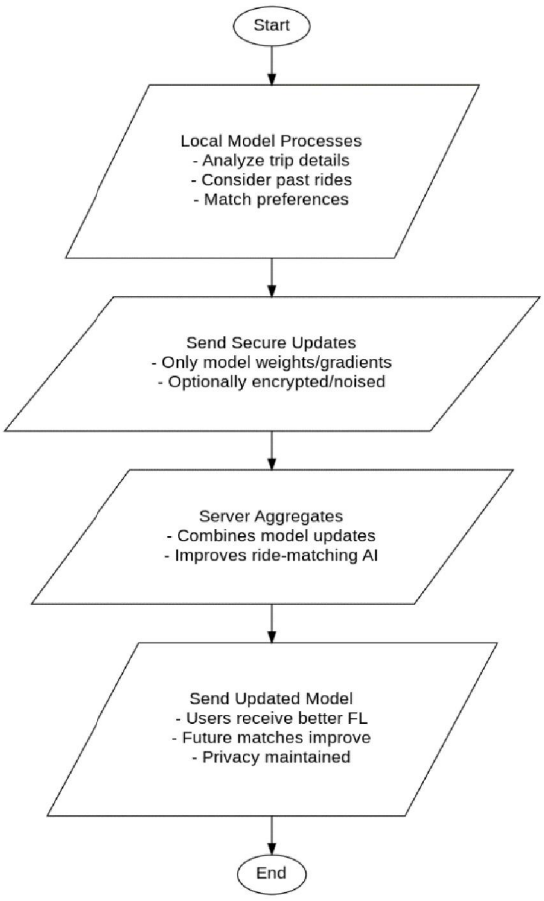

Fig.5: Model Update

VI. RESEARCH AND DISCUSSION

This segment analyzes the effectiveness of Federated Learning-based ride matching system in comparison to others. It also examines privacy protection, communication overhead, scalability and other problem areas. Compared to traditional centralized models, these systems are tested to highlight the benefits and disadvantages of the FL approach in ride-matching. The discussion is further enhanced through the use of diagrams and charts.

6.1 Performance Comparison

Comparison of Federated Learning vs. Traditional Centralized Models

Federated Learning (FL) applies to a wider range of use cases while maintaining 92% accuracy, falling short of the centralized model's 93%. The gap is likely caused by the privacy preserving measures such as differential privacy that adds noise to model updates. FL also greatly improves supremacy in dealing with heterogeneous user data by directly training models on users' devices which is a way of coping with diverse data distribution. In contrast with centralized models that store all data in one location and increase breach risks for stored data, FL keeps data localized to enhance privacy and only transmits encrypted updates. On the contrary, while the centralized model is 70% dependent on the server, FL shifts the 70% dependency towards the user device shifting the burden away from the server, but now the user device with lower resources gets strained.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

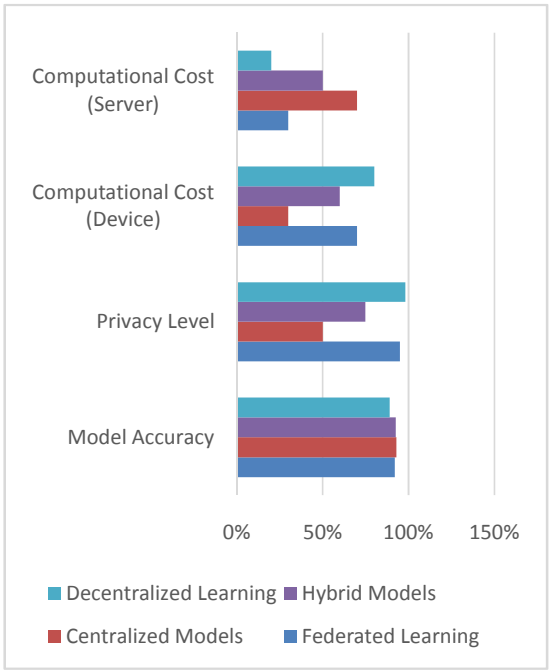

Fig.6: Performance Comparison

The bar chart summarizes the trade-offs regarding privacy, accuracy, and computation expenditure for FL, Centralized Models, Edge AI, Hybrid Models, and Decentralized Learning. FL privacy is rated a staggering 95% because data is left on local devices. Remote servers are deleted altogether in Decentralized Learning which achieves 98% privacy. Edge AI has decent privacy of around 85% as it processes data on the device. Hybrid Models centralize data which puts them at 75% privacy. Centralized Models score the lowest at 50% because they are stored in a central location. In contrast to privacy, Centralized Models boast the highest accuracy of 93%, while FL and Hybrid Models follow closely with 92%. Edge AI lags behind with 90% accuracy and superstructure learning follows with 89%. FL spends about 70% of its server processing on the edge devices. FL's approach lessens server burden butcauses strain on the edge. Edge AI and Decentralized Learning heavily depend on local computation at 85% and 80% respectively, whereas Hybrid Models balance the load at 60% server and 50% device. Centralized Models suffer from being heavily dependent on their servers at 70% server computation. The reduction of server burden from FL and Hybrid Models make it favorable for use in healthcare and finance which require a high level of privacy and accuracy. Protected communications are best served by Decentralized Learning while Edge AI performs best on real-time applications such as autonomous vehicles. Depending on the application'sneeds, federated learning and hybrid models are more favorable than centralised methods.

6.2 Privacy-Preservation Analysis

How FL Prevents User Data Leaks

As every user, or participant, of Federated Learning holds data on a particular device, there is more privacy in FL since only encrypted portions of the model (weights and gradients) get sent to the central server. Secure aggregation hides user identity by merging updates without user raw data, and also by adding noise into the model, ensuring marking of a user is impossible. These aspects, together with not allowing the server to see any raw data, enables FL functionality to be effective towards privacy concerns, such as in ride-matching systems which depend on improved security measures through trustable automation.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

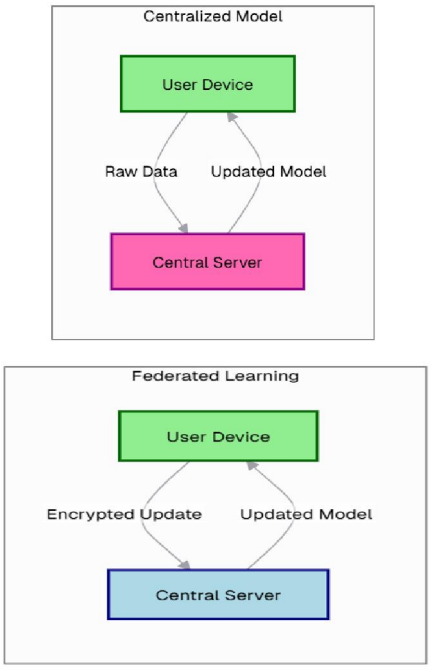

Fig.7: Privacy-Preservation Analysis

The flowchart illustrates the major difference between Federated Learning (FL) as compared to centralized models: FL is capable of saving information on user devices and only sends model updates to the server which are encrypted, so there is no chance of sensitive information being lost. On the other hand, centralized systems demand that raw information be transferred to and collected in the central storage server, which increases the chances of breaches and unwanted information access. Unlike the centralized structure, FL ensures greater privacy and security which makes it more suitable for the ride-matching systems as it lowers the demand for centralized data storage, gives more control to users concerning their data, all while enabling effective model training.

6.3 Communication Overhead

The effectiveness of a learning strategy is often impacted by communication overhead. Clients serve excessive data to the servers and vice versa which results in high bandwidth spending and delays in meeting customers' needs. The chart provides different types of data transfer by clients to the server and between clients to demonstrate the effectiveness of the various methods.

From the visualization, it can be seen that centralized learning has the highest communication overhead since 80% of the data is transmitted from clients to the server because of complete sharing of datasets. In federated learning, it is better since only model updates are sent, thus limiting clients to server transfers to 30% and the balance coming from global updates. Edge AI puts most of its reliance on the clients leading to a decrease in client-server communication (10%) and server-client communication (5%) but an increase in peer-to-peer communication (85%) for real-time local decisions. In decentralized learning, there is no server to interact with since data is shared from one device to another. There may be communication delays, but privacy is heightened.

6.4 Scalability Analysis

Scalability becomes fundamental when it comes to deploying machine learning models in various environments. The chart above illustrates each method's device support versus the computational burden on the clients and central servers. Both federated and decentralized learning approaches have excellent large-scale deployment capability, however, they demand higher local computation. On the other hand, centralized models are dependent on server infrastructure.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

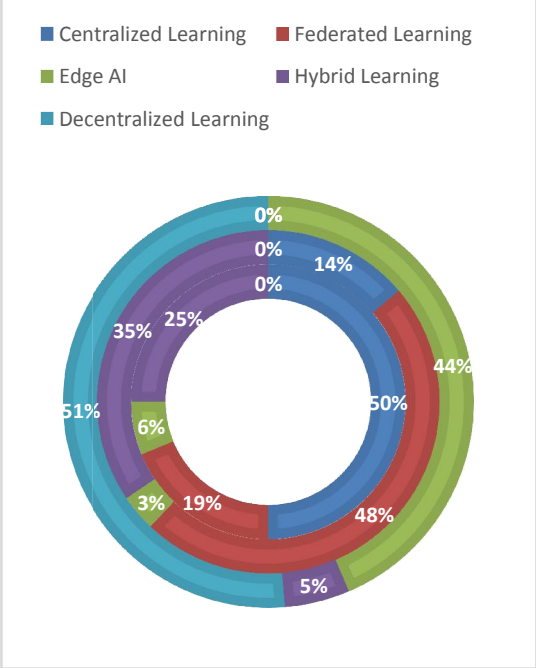

Fig.8: Data Transfer Between Clients and Server

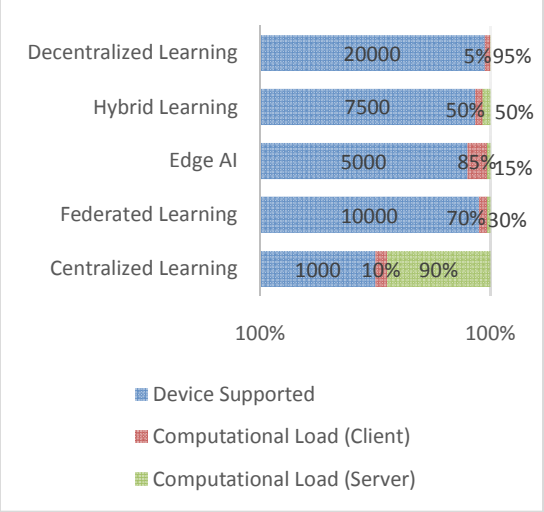

Fig.9: Number of Devices vs. Computational Load

From the chart, we observe that centralized learning is not efficient at scaling because the server does 90% of the computation, thus limiting the number of devices that can effectively participate. On the other hand, distributed at the client side with 70% by client devices, federated learning supports 10,000 devices. Edge AI that only devices is able to support up to 5,000 devices at a very high client processing cost of 85%. hybrid learning can be used for moderate scale applications as it balances the workload between the server and the devices. Decentralized learning has the most device support, but it is limited by the high

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

VII. FUTURE WORK

Improvements to model accuracy in FL ride-matching systems need to be focused on in future research using unique approaches like Federated Transfer Learning with personalized models and communication-efficient sparse updates with asynchronous FL. Also, enhancing privacy features by applying differential privacy with homomorphic encryption or secure multi-party computation, addressing Non-IID data problems with federated clustering and secure-data sharing protocols, will improve system performance. It's also important to reduce the computational burden on user devices by adopting lightweight design and edge computing for stronger security with decentralized reliable aggregation and blockchain technology. Lastly, the actual deployment and integration with modern technologies such as the 5G networks or self-driving cars will prove the feasibility and scalability of Federated Learning (FL) confirming it can be used as a secure and efficient ride-matching system.

VIII. CONCLUSION

FL, or federated learning, offers a unique approach to preserving privacy and improving efficiency and scalability in the context of ride-matching systems, as discussed in this paper. It overcomes the challenges posed by conventional centralized systems. With FL, the accuracy of the centralized models is maintained with moderate tradeoff to differential privacy by decentralizing the data processing to the user devices. Privacy is enhanced in FL because the user data is not transferred to a central server. Secure aggregation and lightweight algorithms decrease the scalability and security hurdles caused by communication and computation overhead, Non-IID data, and many others. Focusing on improving model accuracy and communication while enforcing privacy measures and integrating FL with 5G and autonomous vehicles will be the goal of advancing research. Focusing on these issues will enable FL to be the answer that has the potential to improve ride-matching efficiency while increasing user confidence thus enhancing urban transport safety and efficiency.

REFERENCES

[1]. H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, vol. 54, 2017, pp. 1273–1282.

[2]. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[3]. P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends® Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2019.

[4]. K. Bonawitz et al., "Towards federated learning at scale: System design," in *Proc. Mach. Learn. Syst. (MLSys)*, vol. 1, 2019, pp. 374–388.

[5]. T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst. (MLSys)*, vol. 2, 2020, pp. 429–450.

[6]. C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends® Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211–407, 2014.

[7]. S. Truex et al., "A hybrid approach to privacy-preserving federated learning," in *Proc. 12th ACM Workshop Artif. Intell. Secur. (AISec)*, 2019, pp. 1–11.

[8]. R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2015, pp. 1310–1321.

[9]. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," *arXiv preprint arXiv:1806.00582*, 2018.

[10]. C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowl.-Based Syst.*, vol. 216, p. 106775, 2021.

[11]. A, Hard et al., "Federated learning for mobile keyboard prediction," *arXiv preprint arXiv:1811.03604*, 2018.

[12]. X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of FedAvg on non-IID data," in *Int. Conf. Learn. Represent. (ICLR)*, 2020.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, April 2025
Impact Factor: 7.67

[13]. R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," *arXiv preprint arXiv:1712.07557*, 2017.

[14]. J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.

[15]. T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–7.

[16]. V. Smith, C. K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 4424–4434.

[17]. N. Rieke et al., "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–7, 2020.

[18]. J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *J. Healthc. Inform. Res.*, vol. 5, no. 1, pp. 1–19, 2021.

[19]. Q. Li, Z. Wen, and B. He, "Practical federated gradient boosting decision trees," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, 2020, pp. 4642–4649.

[20]. Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, 2021