RESEARCH PROJECT PROPOSAL

Research Project Title : **Fine-Tuning Large Language Models for Chronic Kidney Disease Prediction**: *A Benchmark Against ML and Deep Learning Models*

1. Project Summary :

Chronic kidney disease (CKD) is a type of kidney disease in which there is a gradual but progressive loss of glomerular filtration rate (GFR) resulting in need for renal replacement therapy, such as dialysis or transplantation. In this particular disease , kidneys slowly get damaged and can't do important jobs like removing waste and keeping blood pressure normal. This disease is a silent killer and is affecting millions of patients worldwide and early detection and prediction of CKD is in high need to slow down the progression of disease and improve patient outcomes. Large Language Models are growing rapidly due to the rise of Transformer architecture introduced by Google DeepMind Researchers and are being implemented in every field of life including Healthcare. LLMs are good at capturing both short-term and long-term dependencies among data which makes them perfect to be implemented on any kind of data. Morever ,similar to Machine Learning and Deep Learning they can be updated with new data to improve their predictive performance . This research aims to evaluate the predictive performance of fine-tuned LLMs while benchmarking them against traditional machine learning and deep learning models.

2. Hypothesis/Basis of Research:

Large Language Models (LLMs) have recently attracted attention in the healthcare domain due to their ability to process complex, context-rich information and predict various diseases while providing evidence-based explanations (Rashed *et al*., 2025 ; Zhou *et al*., 2024). Technology leaders such as Google , Microsoft etc. have introduced medical-specific LLMs (e.g., Med-PaLM 2, BioGPT etc.), recognized for their robust summarization capabilities and their capacity to generate clear, interpretable explanations alongside diagnostic predictions (Google Research, 2023 ; Shool *et al*., 2025 ; Luo *et al*., 2022). Recent work demonstrated that general-purpose LLMs can achieve competitive performance compared to traditional ML and deep learning models on tabular data classification tasks including healthcare datasets (Hegselmann et al., 2023).

This research aims to **benchmark the performance of fine-tuned large language models (LLMs) against traditional ML models** for Chronic Kidney Disease prediction, and to **evaluate their interpretability**. Following the concern raised by Nasarian et al. (2023) about mistrust from “black-box” models, this study will implement a self-explanation prompting approach, enabling the LLM to articulate a natural language reasoning behind each prediction, thereby enhancing transparency and trust without relying on post-hoc methods like SHAP or LIME.

3. Goals/Objectives :

- **Objective (i):** To develop and fine-tune a Large Language Model (LLM) on the local CKD patient dataset for predictive modeling, and to benchmark its performance against established machine learning models. Evaluation will be conducted using accuracy, F1-score, precision, recall, and AUC to determine the most effective model for CKD prediction.

- **Objective (ii):** To enhance the explainability of finetuned model predictions by incorporating *self-explanation prompts* , enabling the model to provide natural language justifications for its decisions. This will help physicians understand the rationale behind each prediction and improve trust in the system's outputs.

- **Identify end user/ beneficiary sector/industry :**

The primary beneficiary of this project is the healthcare sector, particularly professionals and organizations involved in the diagnosis, treatment, and management of kidney diseases. Specifically, this research is being conducted in collaboration with, and with the support of, the Pakistan Kidney Center, Khokhar Maira, Abbottabad, Khyber Pakhtunkhwa, Pakistan.

4. Introduction :

Chronic kidney disease (CKD) is a type of kidney disease in which there is a gradual but progressive loss of glomerular filtration rate (GFR) over a period of more than 3 months (Stevens et al ,2013) . The World Health Organization (WHO) states that chronic kidney disease (CKD) is a major contributor to mortality and impaired functioning on a global scale (WHO, 2005).It is a silent killer as there are no physical symptoms in the early stage. Around 10% of the world's population is believed to have CKD, which results in millions of deaths each year due to the disease or its associated complications. There are different stages of CKD and mortality rate varies depending on the severity and stage of the disease and the presence of other underlying health conditions (Mahadevan, 2019). Individuals in the advanced stages of CKD may need dialysis or a kidney transplant to stay alive when their kidneys can no longer perform their functions adequately. The mortality rate for these patients can be high, especially for those who do not receive timely treatment or who have other serious health problems. It is important to note that early detection and management of CKD can help to slow the progression of the disease and reduce the risk of serious complications and death (Aljaaf et al., 2018). This highlights the importance of regular check-ups and screening for CKD, especially for those at increased risk, such as people with diabetes, high blood pressure, or a family history of kidney disease.

CKD is a growing global health concern, and early prediction of the disease is critical for effective management and treatment (Celik et al., 2016; Mahadevan, 2019).). Traditional diagnostic methods such as blood and urine tests, while clinically standard, are often time-consuming, costly, and limited in their predictive scope. In recent years, machine learning (ML) has emerged as a promising alternative, offering scalable and data-driven solutions for CKD risk assessment. Machine learning is a branch of AI that allows computers to learn from historical data and make predictions. ML algorithms including logistic regression, decision trees, support vector machines (SVM), random forests, and artificial neural networks (ANNs) have demonstrated strong performance in identifying CKD risk factors and predicting disease onset using structured data from medical records, laboratory tests, and demographic profiles (Silveira et al., 2022; Chittora et al., 2021; Bai et al., 2022).

Despite their success, traditional ML models often struggle with capturing complex feature interactions and providing solid reasoning behind predictions. This is where Large Language Models (LLMs) come into action. LLMs are transformer-based models designed to capture complex relationships across input features, making them inherently strong at reasoning (Wang et al., 2021). Their ability to model both short and long-range dependencies allows them to infer nuanced interactions such as how hemoglobin levels, age, and proteinuria jointly influence CKD risk. This reasoning capability is crucial in clinical settings, where healthcare professionals require not just predictions but also interpretable explanations to guide medical decisions and treatment planning (Stevens et al., 2013; Couser et al., 2011).

In this work , we proposed to use fine-tuned LLMs on CKD-specific tabular data to evaluate their effectiveness in predicting disease risk and providing clinically meaningful rationales.

5. Research plan/ Methodology :

1. <u>Data Collection:</u>

The first step in this study will be to collect a large and representative dataset of medical records of individuals with and without CKD from Pakistan Kidney Center, Abbottabad. This dataset will include the following components:

- **Demographic Information**: Age, gender, ethnicity, etc.
- **Lab Results**: Blood test reports (e.g., creatinine, eGFR levels).
- **Clinical Measurements**: Symptoms and physical examinations (e.g., blood pressure, diabetes status).

2. <u>Data Preparation:</u>

The collected data will then be cleaned and pre-processed to ensure its quality and consistency. This will include following steps:

- **Data Cleaning**: Handle missing values, outliers, and errors in the dataset.
- **Feature Selection**: Identify important attributes (features) that influence CKD prediction via statistical methods , correlation analysis etc.
- **Data Splitting:** Partitioning into training, validation, and test sets (e.g., 70/15/15).

3. **Data Serialization :**

To enable the use of Large Language Models (LLMs) for CKD prediction, the structured tabular data must be transformed into serialized formats that LLMs can interpret effectively. This involves converting each patient record into a natural language description that captures all relevant features in a coherent and context-rich manner. For example:

"*Patient is a 55-year-old male with blood pressure of 140/90 mmHg, … , He is diagnosed with diabetes and has chronic kidney disease*."

Other approaches such as **HTML tables**, **JSON objects**, and **YAML** formats can also be considered for Tabular Data Serialization.

4. **Model Training and Testing:**

Once the CKD dataset has been serialized into LLM-compatible formats, the next step is to leverage pre-trained Large Language Models (LLMs) for predictive modeling. This step includes:

Pre-trained Model Selection :

A suitable open-source LLM will be selected based on compatibility with tabular reasoning and available computational resources.

Fine-Tuning Strategy :

The selected model will be fine-tuned using the serialized CKD dataset. We will explore **Parameter-Efficient Fine-Tuning (PEFT)** methods such as **LoRA**, **QLoRA**, and other lightweight adaptation techniques to efficiently adapt the model to CKD-specific patterns.

5. **Model Validation and Evaluation:**

Baseline results of traditional ML/DL models (e.g., SVM, Random Forest, Decision Tree, Neural Networks) will be summarized from peer-reviewed CKD studies, typically reported using :

- **Accuracy**
- **precision**
- **recall**
- **F1-score**
- **AUC-ROC** etc.

The fine-tuned LLM will be evaluated independently on the same metrics to assess reliability. Beyond performance, the study will also test the LLM's ability to provide **clinically meaningful explanations** for its predictions—an aspect where conventional ML models act as black boxes.

6. **Model Deployment:**

After benchmarking, the **best-performing model** will be selected based on evaluation metrics such as **accuracy, precision, recall, and F1-score**, along with its ability to provide **interpretable outputs**.

- If the **fine-tuned LLM** outperforms traditional ML models, it will be adopted directly for deployment.
- If not, a **hybrid solution** will be designed: a traditional ML model will handle **prediction tasks**, while the fine-tuned LLM (trained on the same dataset) will generate **interpretability and reasoning** behind each prediction.

Below is the visual diagram showing how the data would be fed to the LLM and how the prediction would be obtained from it .

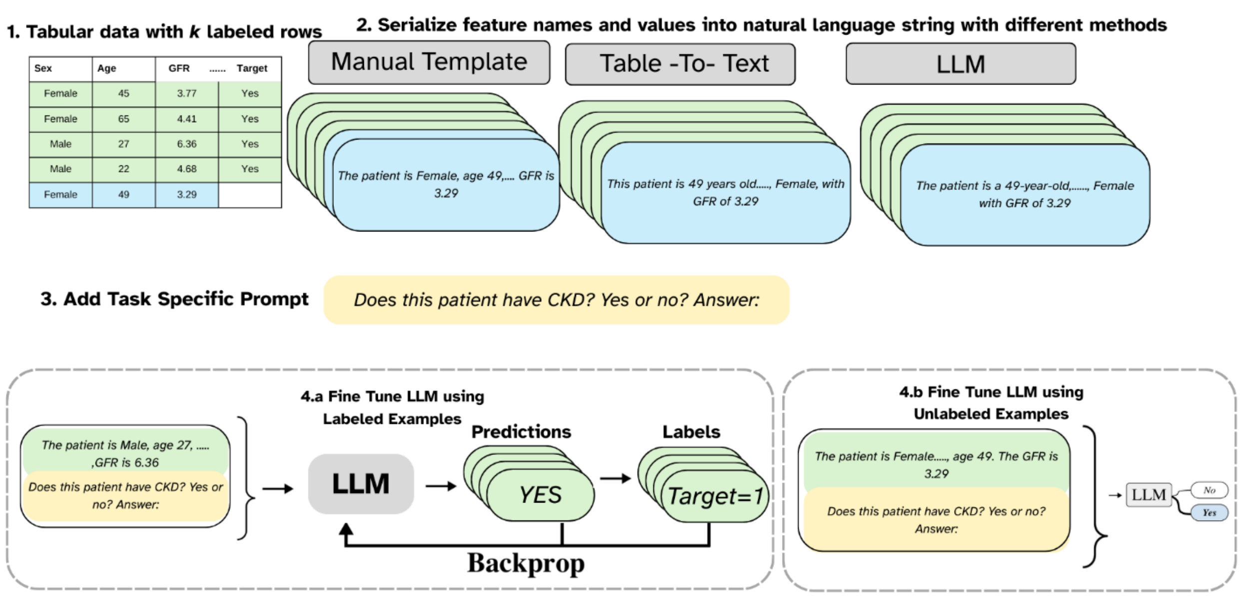

Figure i: Tabular Data Serialization and LLM Fine-Tuning Workflow

The attached visual representation below, maps out the end-to-end workflow of training the model on serialized CKD data and getting interpretable predictions during inference from the LLM once the model is trained .

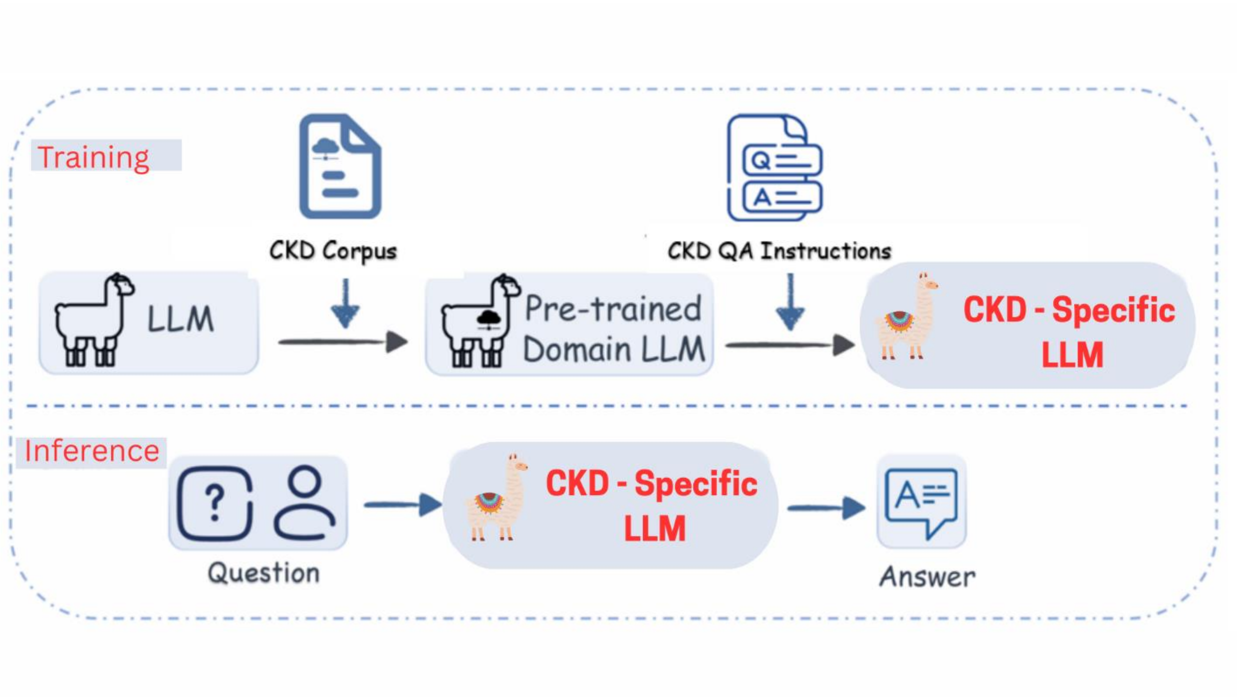

Figure ii: End to End Workflow of CKD-specific LLM Model

This final model will be deployed at the **Pakistan Kidney Center** to provide accurate CKD risk prediction along with transparent explanations, supporting healthcare professionals in clinical decision-making. In parallel, a **user-friendly mobile and web application** will be developed, allowing individuals to input their health parameters (e.g., lab results, vitals) and receive instant CKD risk assessments. The application will not only deliver predictions but also provide clear reasoning behind the outputs, empowering users to better monitor and manage their kidney health independently.

5B. Ethical Consideration:

This study will be conducted at PAF-IAST and Pakistan Kidney Center, KPK, Pakistan, in compliance with the ethical committee guidelines. All experimental methods will be performed in accordance with the specified checklist and will be approved by the Research Ethics Committee (REC) of PAF-IAST prior to the commencement of the study. The dataset to be utilized in this study will be collected from PKC, with informed consent and assent obtained from all participants. Due to privacy and ethical concerns, the datasets to be used in this study are not currently available for public access.

<u>Proposed Work Plan:</u>

TASKS	SEP 2025	OCT 2025	NOV 2025	DEC 2025	JAN 2026	FEB 2026	MAR 2026	APRIL 2026	MAY 2026	JUN 2026	JULY 2026	AUG 2026
Research Area												
Literature review												
Data Collection												
Experimental work												
Data Analysis												
Result and Discussion												
Write-up												
Product												
Paper Publication												

6. References :

1. Aljaaf, A. J., Al-Jumeily, D., Haglan, H. M., Alloghani, M., Baker, T., Hussain, A. J., & Mustafina, J. (2018). Early prediction of chronic kidney disease using machine learning supported by predictive analytics. *2018 IEEE Congress on Evolutionary Computation (CEC)*, 1–9.
2. Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific Reports, 12*(1), 1–8.
3. Celik, E., Atalay, M., & Kondiloglu, A. (2016). The diagnosis and estimate of chronic kidney disease using the machine learning methods. *International Journal of Intelligent Systems and Applications in Engineering, 4*(Special Issue-1), 27–31.

4. Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., Jasiński, M., Jasiński, Ł., Gono, R., & Jasińska, E. (2021). Prediction of chronic kidney disease—a machine learning perspective. *IEEE Access, 9*, 17312–17334.
5. Couser, W. G., Remuzzi, G., Mendis, S., & Tonelli, M. (2011). The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney International, 80*(12), 1258–1270. https://doi.org/10.1038/ki.2011.368
6. Google Research. (2023). *Med-PaLM 2: Towards expert-level medical question answering with LLMs.* Retrieved from https://research.google
7. Hegselmann, S., Buendia, A., Glaese, A., Li, X. L., Rutherford, E., Stammers, T., & McAleese, N. (2023). TabLLM: Few-shot classification of tabular data with large language models. *arXiv preprint arXiv:2305.02920*.
8. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T. Y. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics, 23*(6), bbac409. https://doi.org/10.1093/bib/bbac409
9. Mahadevan, V. (2019). Anatomy of the kidney and ureter. *Surgery (Oxford), 37*(7), 359–364.
10. Nasarian, E., Gholipour, K., Salimi, A., Ahmadi, E., & Gandomi, A. H. (2023). Explainable artificial intelligence for healthcare: From black box to glass box. *Information Fusion, 90*, 364–399. https://doi.org/10.1016/j.inffus.2022.09.015
11. Rashed, A., Abdel-Basset, M., & Mohamed, R. (2025). Large language models in healthcare: Opportunities, challenges, and future directions. *Healthcare Analytics, 7*(1), 100259.
12. Shool, N., Gupta, A., & Singh, R. (2025). Medical large language models: Applications and challenges in clinical decision support. *Journal of Medical Systems, 49*(2), 45.
13. Silveira, A. C. M. da, Sobrinho, Á., Silva, L. D. da, Costa, E. de B., Pinheiro, M. E., & Perkusich, A. (2022). Exploring early prediction of chronic kidney disease using machine learning algorithms for small and imbalanced datasets. *Applied Sciences, 12*(7), 3673.
14. Stevens, P. E., Levin, A., & Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. (2013). Evaluation and management of chronic kidney disease: Synopsis of the KDIGO 2012 clinical practice guideline. *Annals of Internal Medicine, 158*(11), 825–830. https://doi.org/10.7326/0003-4819-158-11-201306040-00007
15. Wang, A., Cho, K., & Lewis, M. (2021). Asking the right questions: Large language models for reasoning over tabular data. *arXiv preprint arXiv:2109.02701*.
16. World Health Organization. (2005). *Preventing chronic diseases: A vital investment.* Geneva: WHO.
17. Zhou, Y., Zhang, H., Wang, J., & Xu, X. (2024). Large language models in clinical research: Opportunities and challenges. *Frontiers in Artificial Intelligence, 7*, 1384. https://doi.org/10.3389/frai.2024.01384

7. Impact in Quantifiable Terms (Impact of proposed research on teaching/training of manpower, institutional capacity building and on local industry; on economic development of national, regional and community development).

The proposed research on Chronic Kidney Disease (CKD) prediction using Finetuned Large Language Model has the potential to significantly impact the teaching/training of manpower, institutional capacity building, and local industry, as well as the economic development of the national, regional, and community levels.

A. Impact on Teaching/Training of Manpower :

This research will demonstrate the practical application of Machine Learning and Advanced AI specifically LLMs in clinical diagnosis , which can lead to the development of Tech Departments at HealthCare Organizations specifically for integrating AI for Medical Diagnosis of various diseases. Morever , this can lead various institutions and healthcare organizations introduce workshops and certification courses or medical practitioners and IT professionals, focusing on the practical implementation of AI in healthcare, potentially certifying **hundreds of professionals annually** in these advanced skills.

B. Impact on Institutional Capacity Building :

1.Strengthening of healthcare infrastructure:

The research can contribute to the strengthening of healthcare infrastructure (Hospitals, Clinics, Laboratories, and Institutions) in Khyber Pakhtunkhwa, Pakistan, by providing evidence-based support for the adoption of machine learning-based CKD prediction tools in clinical settings. This can lead to improved patient care and substantially reduced healthcare costs.

2.Establishment of research centers:

The research can facilitate the establishment of research centers focused on machine learning and advanced ai applications in healthcare, particularly in the area of CKD prediction. These centers can conduct further research, develop new machine learning algorithms, and provide training to healthcare professionals.

C. Impact on Local Industry :

1. **Growth of HealthTech Startups**

Successful deployment of the fine-tuned LLM at Pakistan Kidney Center and the development of a user-friendly mobile/web application will inspire and enable local startups to innovate in the HealthTech sector, particularly in areas of predictive analytics, remote patient monitoring, and AI-driven diagnostics for various diseases. This could lead to the emergence of new HealthTech startups focused on AI solutions for healthcare in the region within 3–5 years.

2. **Increased Demand for AI / ML Talent**
As local industries adopt AI solutions, there will be a surge in demand for AI/ML engineers, data scientists, and healthcare IT specialists, stimulating job creation and economic growth in the tech sector.

D. Economic Development:

1.Reduced Healthcare Costs

Early detection of CKD can significantly reduce treatment costs by identifying the disease at its initial stages. Patients would save substantial amounts of money otherwise spent on dialysis or transplants. Reports indicate that improved early detection could save Pakistan **tens of millions of dollars annually**, easing the burden on both families and the healthcare system.

2.Improved Public Health Outcomes

Early CKD prediction can help lower the mortality rate associated with the disease. By enabling timely interventions, patients' chances of survival and quality of life improve. Advancement in such research fields can make Pakistan a **leader in cutting-edge medical AI** and enhances its global reputation in health research.

3. Enhanced Productivity

Reducing CKD prevalence will improve the overall health of Pakistani citizens, allowing them to contribute more productively to society and the economy. A healthier population directly supports national development, while innovations in healthcare can attract investment, foster local health-tech industries, and generate **revenue streams** for the country.

Overall, the proposed research on Chronic Kidney Disease (CKD) prediction using advanced AI has the potential to make a significant contribution to the teaching/training of manpower, institutional capacity building, local industry, and economic development in Khyber Pakhtunkhwa, Pakistan, and beyond.

8. Provincial Development Strategies (How and which of the strategies will be addressed in this study? Justify how the proposed research will contribute to achieve the development strategies of Khyber Pakhtunkhwa)

The proposed research on Chronic Kidney Disease (CKD) prediction using fine-tuned Large Language Models aligns with several of Khyber Pakhtunkhwa's development strategies and has the potential to contribute significantly to achieving these goals. Here's how:

1. Strengthening Healthcare Infrastructure and Service Delivery:

The development of an AI-powered CKD prediction system with interpretable explanations directly addresses the provincial government's objective of strengthening healthcare infrastructure and service delivery. Early detection and prediction of CKD through fine-tuned LLMs can lead to more effective treatment and management, reducing the burden on healthcare facilities like the Pakistan Kidney Center and improving patient outcomes. The deployment of both clinical decision-support tools and user-friendly mobile applications will enhance service accessibility across urban and rural areas of KPK.

2.Enhancing Research and Innovation Capacity :

This research demonstrates practical LLM implementation in clinical diagnosis, directly supporting KP's strategy of enhancing research and innovation capacity. The successful CKD prediction project will establish KP as a regional leader in healthcare AI research, attracting academic partnerships, research funding, and innovation investments. This positions the province at the forefront of cutting-edge medical AI applications and contributes to building a robust knowledge economy.

3. Job Creation in AI and Healthcare Technology :

The successful deployment of this project will cause the establishment of specialized AI and healthcare technology departments within existing healthcare organizations and stimulate the growth of new tech enterprises focused on medical AI applications which will inspire and enable local startups to innovate in the HealthTech sector, particularly in areas of predictive analytics, remote patient monitoring, and AI-driven diagnostics for various diseases. This will significantly increase demand for skilled AI/ML professionals across the province, creating high-value employment opportunities that align with KP's vision of building a knowledge-based economy.

4. Knowledge Economy Building :

This advanced AI project will serve as a flagship demonstration of practical AI implementation in healthcare, drawing attention from educational institutions and healthcare organizations to the transformative potential of AI in medical applications. The successful implementation will encourage these organizations to introduce diverse workshops and certification programs in healthcare AI, which could collectively train **thousands of students and professionals** in this emerging field. This mass skilling initiative directly contributes to making KP's citizens more technically proficient and competitive in the digital economy.

5. Promoting Sustainable Development and Environmental Protection:

The research indirectly contributes to the province's goal of promoting sustainable development and environmental protection by reducing the environmental footprint of healthcare services. Early detection and management of CKD can prevent the need for dialysis, which is a resource-intensive and environmentally taxing procedure.

6. Empowering Communities and Promoting Social Inclusion:

The research supports the province's efforts to empower communities and promote social inclusion by addressing a disease that disproportionately affects marginalized populations. Early detection and prediction can improve access to healthcare services for vulnerable communities and reduce the socioeconomic impact of CKD.

In summary, the proposed research on CKD prediction using advance AI aligns with Khyber Pakhtunkhwa's development strategies by strengthening healthcare infrastructure, promoting human capital development, enhancing research and innovation, promoting sustainable development, and empowering communities. By addressing a critical healthcare challenge, the research has the potential to significantly improve the lives of individuals affected by CKD and contribute to the overall development of the province.

9. Collaborating organization :

Pakistan Kidney Center Khokhar Maira, Abbottabad, Khyber Pakhtunkhwa, Pakistan