Contents

[Project Title: Securing Interpretable Deep Learning Systems: Adversarial Threats, Stealthy Attacks, and Defense Mechanisms 1](#_Toc194951557)

[Project Title: Comprehensive Evaluation of Adversarial Robustness in Deep Learning: Architecture, Diversity, and Defense Analysis 3](#_Toc194951558)

[Project title: Explainable Artificial Intelligence for Trustworthy and Transparent Decision-Making in Medical Applications 5](#_Toc194951559)

[Project Title: Robust Malware Detection in Adversarial Environments: Analysis, Evaluation, and Defense Strategies 6](#_Toc194951560)

[Project Title: Behavioral Biometrics for Continuous and Adversarially Robust User Authentication on Smartphones 9](#_Toc194951561)

[Project Title: Multimodal, Explainable, and Adversarially-Robust Deep Learning for Alzheimer’s Disease Progression Detection 11](#_Toc194951562)

[Project Title: Explainable Dynamic Ensemble Learning with Late Fusion of Multimodal Data for Intelligent Decision Support 13](#_Toc194951563)

[Project Title: Explainable and Dynamic Ensemble Models for ICU Mortality and Length-of-Stay Prediction 15](#_Toc194951564)

# Project Title: Securing Interpretable Deep Learning Systems: Adversarial Threats, Stealthy Attacks, and Defense Mechanisms

**[Image1]**

**Project Description:** The growing integration of deep learning (DL) models into high-stakes domains, such as healthcare, finance, and autonomous systems, has made interpretability a cornerstone of trustworthy AI. Interpretable Deep Learning Systems (IDLSes), which combine powerful neural networks with interpretation models, aim to provide transparency into the decision-making process. However, the assumption that interpretation inherently adds security has recently been challenged.

At InfoLab, Sungkyunkwan University (SKKU), our research group has pioneered a comprehensive investigation into the security vulnerabilities of IDLSes, revealing how interpretation mechanisms themselves can be exploited by adversarial actors. Through a coordinated body of work encompassing three major research efforts, we have introduced novel attack techniques and analyzed defense strategies, reshaping the understanding of what it means for a deep learning system to be explainable and secure.

**Key Contributions from InfoLab, SKKU:**

1. **SingleADV: Single-Class Target-Specific Attack**We introduced SingleADV, a universal perturbation-based adversarial attack that targets a specific object class, misclassifying it into a chosen target category. Crucially, SingleADV preserves interpretation maps that resemble benign ones, making it difficult for both automated systems and human observers to detect the manipulation. This attack works effectively in both white-box and black-box settings across multiple model and interpreter architectures.
2. **AdvEdge and AdvEdge+:** Interpretation-Guided Stealthy Attacks  
   These attacks exploit the edge-sensitive regions of input images, identified by interpretation maps, and inject subtle perturbations aligned with those areas. This strategy enhances stealth and minimizes detection, making the adversarial samples nearly indistinguishable from benign inputs. The methods have shown strong resilience and cross-model transferability, and can be applied across different datasets and interpreter types (e.g., CAM, Grad, MASK).

**[Image2]**

1. **QuScore: Stealthy Query-Efficient Opaque Attack**Recognizing real-world black-box constraints, we developed QuScore, a query-efficient and model-agnostic attack tailored for black box settings, where internal model access is not available. Leveraging transfer-based strategies and a custom genetic algorithm, QuScore maintains interpretation fidelity while achieving over 95% attack success rate with a low query budget. It demonstrates that even limited feedback can be enough to compromise both prediction and interpretation.

**Project Objectives:**

* Systematically study and break the assumptions of interpretability-driven robustness in DL systems.
* Design category-specific, stealthy, and transferable adversarial attacks that target both prediction outputs and interpretation layers.
* Explore the vulnerabilities of IDLSes in real-world black-box and query-limited settings.
* Develop robust interpretation-aware defenses, including adversarial training and ensemble-based detection strategies.
* Provide open-source tools and benchmarks to facilitate future research in adversarial robustness and trustworthy AI.

**Research Impact:** Our research at InfoLab (SKKU) reveals that interpretability is not inherently secure, and in fact, can serve as an attack surface when exploited thoughtfully. These findings are crucial for developers, regulators, and researchers designing DL systems in safety-critical environments. By unveiling these hidden threats and proposing initial defenses, we contribute toward building the next generation of resilient, trustworthy, and explainable AI systems.

# Project Title: Comprehensive Evaluation of Adversarial Robustness in Deep Learning: Architecture, Diversity, and Defense Analysis

**Project Description:** Adversarial attacks pose a serious challenge to the reliability and security of deep learning (DL) models. These attacks, often crafted by introducing imperceptible perturbations to input data, can cause models to make incorrect predictions with high confidence. As a result, understanding and mitigating such threats has become a critical area of research in the field of trustworthy AI. Defenses against adversarial attacks range from input preprocessing and adversarial training to robust model design, yet no single approach has proven universally effective. At **InfoLab, Sungkyunkwan University (SKKU)**, our research group undertakes a systematic investigation into the **effectiveness of various adversarial attacks and corresponding defense mechanisms**, providing a comprehensive analysis of how deep learning models respond to different threat vectors across multiple architectures and datasets.

**[image3]**

**Core Research Themes and Contributions**

**1. Multi-Dimensional Analysis of Adversarial Attacks and Defenses**

Through large-scale experimentation, we investigated how factors such as model complexity, architectural diversity, and training datasets impact robustness against adversarial attacks. We evaluated several white-box and black-box attack methods (e.g., FGSM, PGD, C&W, SimBA, HopSkipJump, MGAAttack, and Boundary Attack) across popular model families (e.g., VGG, ResNet, DenseNet, MobileNet, Inception, Xception, ShuffleNet) and datasets (ImageNet, CIFAR-10, CIFAR-100). Key findings include:

* **Deeper models** (with more layers) tend to be more robust in general, requiring higher perturbation magnitudes and longer attack times.
* **Model parameter count** alone does not determine robustness; architecture design plays a more critical role.
* **Black-box attacks** become increasingly ineffective as model complexity grows, while simple preprocessing-based defenses (e.g., JPEG filtering, median smoothing) can reduce success rates significantly.

**2. Impact of Architectural Variations on Robustness**

We performed a controlled analysis on variations of popular DL architectures, including:

* **VGG models with and without Batch Normalization (BN)**: Although BN improves accuracy, it may increase vulnerability to adversarial attacks by enabling gradient masking and reducing feature diversity.
* **Inception variants**: Inception ResNet V2 and Inception V4 showed increased resistance compared to Inception V3, due to architectural enhancements such as residual blocks and bottleneck modules.
* **MobileNet family**: MobileNet V3 (both small and large) offered better robustness than V2, credited to the integration of squeeze-and-excitation (SE) modules and hard-swish activations.

**3. Behavior of Attacks Across Datasets and Defensive Strategies**

We highlighted how dataset characteristics (e.g., resolution, number of classes) affect model behavior under attack. Notably:

* Attacks on models trained on high-resolution data required more sophisticated perturbations than those on low-resolution datasets.
* Models trained on datasets with more classes but smaller image sizes displayed inconsistent robustness patterns compared to other models.
* **Preprocessing defenses**, while simple, proved surprisingly effective against black-box attacks when tuned correctly.

**Project Objectives**

* Provide an empirical foundation for understanding the **robustness landscape** of DL models against adversarial threats.
* Evaluate the impact of **architectural decisions and model upgrades** on adversarial vulnerability.
* Develop practical guidelines for deploying **robust and defensible deep learning models** in real-world environments.
* Investigate lightweight and **efficient defense mechanisms** that do not require architectural retraining.

**Research Impact:** This project provides a systematic and reproducible benchmark for adversarial robustness analysis, offering insights for both researchers and practitioners seeking to deploy secure AI solutions. By exposing previously underexplored relationships between model design, training data, and adversarial behavior, InfoLab at SKKU is helping shape the future of **trustworthy and resilient AI systems**.

# Project title: Explainable Artificial Intelligence for Trustworthy and Transparent Decision-Making in Medical Applications

**Project Description:** The project seeks to address the growing need for transparency, accountability, and interpretability in artificial intelligence (AI) systems used in healthcare. As deep learning and other machine learning techniques become integral to medical diagnostics, prognosis, and treatment planning, the "black-box" nature of many AI models poses significant challenges for clinical adoption, regulatory approval, and patient trust.

This research focuses on developing and evaluating **Explainable Artificial Intelligence (XAI)** methodologies tailored to medical contexts. The goal is to ensure that AI-driven decisions are not only accurate but also interpretable by clinicians, understandable to stakeholders, and compliant with ethical and legal standards.

The project adopts a **multi-dimensional framework** for explainability, structured around four core axes:

1. **Data Explainability** – Enhancing the transparency of input features and their contributions to model decisions.
2. **Model Explainability** – Designing or adapting models that are inherently interpretable or hybrid approaches that balance accuracy with clarity.
3. **Post-hoc Explainability** – Applying state-of-the-art interpretability techniques such as SHAP, Grad-CAM, LIME, and Layer-wise Relevance Propagation to make predictions understandable.
4. **Assessment of Explanations** – Developing robust evaluation and human-centered evaluation strategies to measure explanation quality, usability, and trustworthiness.

By bridging the gap between algorithmic complexity and clinical insight, this project aims to empower healthcare professionals with AI systems that are **transparent, reliable, and ethically grounded**. Additionally, the project examines explainability from multiple perspectives—technical, clinical, and regulatory—to inform the development of AI models that align with real-world medical decision-making requirements.

By doing so, this research contributes to the foundation of trustworthy AI in medicine, promoting the safer and more responsible use of intelligent systems in healthcare environments.

# Project Title: Robust Malware Detection in Adversarial Environments: Analysis, Evaluation, and Defense Strategies

**Project Description:** The dynamic evolution of malware, combined with increasingly sophisticated evasion techniques such as packing, obfuscation, and polymorphism, presents a significant challenge to conventional security mechanisms. As a result, **machine learning (ML)-based malware detection** systems are being adopted widely due to their ability to generalize and automate malware identification. However, these systems are also **susceptible to adversarial threats**, and current detection solutions struggle with robustly identifying evasive or morphed malware.

To address this critical issue, **InfoLab at Sungkyunkwan University (SKKU)** has led a comprehensive research project that spans **three key investigations**, each targeting a unique vulnerability in ML-based malware detection pipelines—from data representation and feature manipulation to evasion through software packing.

**Core Research Contributions**

**1. Spectral Analysis of Control Flow Graphs for Malware Detection**

In this direction, we propose a novel approach for malware classification using **spectral representations** of control flow graphs (CFGs). Leveraging **heat and wave kernels**, the research extracts size- and permutation-invariant graph signatures for malware detection. Applied to a dataset of over 37,000 Windows executables, and evaluated across eight ML models, the method achieves an accuracy of **up to 95.9%**.  
*Key Insight:* Spectral signatures provide a scalable and effective alternative to byte-level feature extraction, particularly in adversarial scenarios where malware structures are manipulated.

**2. MLxPack: Investigating the Effects of Packers on ML-Based Malware Detection**

This work examines how **packing techniques**—used to disguise malicious intent—affect the accuracy of ML classifiers. Using a large-scale dataset of **107,000 packed and unpacked samples** and analyzing both **static and dynamic features**, the research shows that **dynamic analysis significantly improves detection robustness**. The study also demonstrates that hybrid (static + dynamic) feature sets improve classifier performance and that **packed malware is harder to detect without robust multi-dimensional analysis**.  
*Key Insight:* Detection systems must account for packing effects through diverse feature representations and combine multiple analysis perspectives.

**3. Visualization-Based Malware Analysis Using Feature Fusion**

Focusing on **Android malware**, this study introduces a feature fusion technique that combines **handcrafted texture descriptors** (e.g., GIST, LBP, GLCM) with **deep CNN features** extracted from grayscale images of malware components (e.g., classes.dex, manifest files). Evaluated on the DREBIN dataset across various machine learning models, the fusion-based approach achieves a classification accuracy of **93.24%**.  
*Key Insight:* Visualization-based static analysis provides a powerful means to detect Android malware while mitigating the effects of code obfuscation and packing.

**Project Objectives**

* Develop robust, interpretable, and adversarial-resilient malware detection pipelines.
* Analyze the **impact of binary obfuscation and packing techniques** on static and dynamic ML-based detection models.
* Investigate **multi-view and feature-fusion strategies** to improve classification accuracy under adversarial transformations.
* Design efficient, scalable representations (e.g., spectral graph embeddings) to capture malware behavior invariant to structural changes.
* Provide **benchmark datasets, evaluation protocols**, and reproducible experiments for the research community.

**Research Impact:** This project by InfoLab at SKKU represents a **multidimensional approach to adversarial malware analysis, bridging the gap between machine learning** robustness and real-world evasive tactics. The findings contribute actionable insights into:

* **Designing detection models that withstand advanced evasion techniques** like packing and morphing,
* **Developing interpretable and generalizable malware representations**, and
* **Guiding future defense frameworks** in both desktop and mobile environments.

Together, these efforts establish a strong foundation for **next-generation malware detection systems** that are secure, explainable, and resilient against adversarial manipulation.

# Project Title: Behavioral Biometrics for Continuous and Adversarially Robust User Authentication on Smartphones

**Project Description:** Traditional authentication methods—such as passwords, PINs, and even biometric systems (fingerprint, facial recognition)—typically secure mobile devices only at the point of entry. However, they fail to offer protection throughout a session, leaving devices vulnerable to unauthorized access when unattended. To bridge this security gap, the research group **InfoLab at Sungkyunkwan University (SKKU)** has led a series of studies on **continuous, sensor-based, and adversarially-aware user authentication mechanisms**.

This project focuses on designing and evaluating **practical, deep-learning-powered implicit authentication systems** using **motion and touch sensor data** collected unobtrusively from smartphones. It addresses real-world deployment constraints, robustness against adversarial conditions, and usability trade-offs to build trustworthy systems that authenticate users continuously based on their behavioral patterns.

**Core Contributions and Insights**

**1. MotionID: Toward Practical Behavioral-Based Implicit Authentication**

MotionID introduces a **comprehensive continuous authentication framework** based on touch and motion sensor data (accelerometer, gyroscope, magnetometer, elevation, gravity). Unlike previous solutions that assume static sampling rates or constrained environments, MotionID:

* Dynamically adapts to **device-specific sensor variations** using a “Global Mobile Average” approach.
* Supports **cross-app and unconstrained usage scenarios** (not limited to specific apps or tasks).
* Operates with **short-duration samples** (1–5 seconds) to rapidly identify users in real-world settings.
* Achieves **F1-scores up to 98.5%** while maintaining low power and memory usage, making it highly deployable.

**2. AUToSen: Deep-Learning-Based Continuous Authentication**

AUToSen is a **high-frequency deep learning authentication system** that uses built-in smartphone sensors (accelerometer, gyroscope, magnetometer) to model users’ behavioral traits.

* Utilizes LSTM-based models to handle time-series sensor data.
* Demonstrates **state-of-the-art accuracy**, with F1-scores of ~98%, equal error rates (EER) as low as **0.09%**, and authentication delays as short as 0.5 seconds.
* Operates effectively with **minimal sensor inputs**, validating the feasibility of lightweight authentication without user inconvenience or privacy intrusion.

**Research Objectives**

* Design and evaluate **implicit, behavioral-based authentication mechanisms** that work in real time on commodity smartphones.
* Overcome **practical limitations** such as device diversity, sensor sampling inconsistencies, and unconstrained user behavior.
* Investigate how to **balance robustness and usability**, including power efficiency and authentication delay.
* Ensure authentication systems remain resilient in adversarial environments where attackers may attempt to mimic behavior.
* Provide a **scalable framework** that supports deployment across heterogeneous devices and user populations.

**Research Impact:** This project offers a significant advancement in **mobile security** by transforming passive sensor data into powerful behavioral signatures for **continuous, transparent, and secure user authentication**. The solutions developed by InfoLab (SKKU) represent a significant step toward real-world, adversarially aware**, and privacy-preserving authentication systems**, with strong implications for mobile banking, healthcare, and sensitive enterprise environments.

------------------------------------------Medical Research Project ------------------------------------

# Project Title: Multimodal, Explainable, and Adversarially-Robust Deep Learning for Alzheimer’s Disease Progression Detection

**Project Description:** Alzheimer’s disease (AD) is a progressive, neurodegenerative disorder with no known cure. Accurate early prediction of its progression—from cognitively normal (CN) and mild cognitive impairment (MCI) to full AD—is vital for enabling timely interventions. However, traditional diagnostic models based on a single data modality or one-time observations are inadequate due to the disease’s heterogeneous and longitudinal nature.

To address these challenges, **InfoLab at Sungkyunkwan University (SKKU)** has spearheaded a research initiative that spans **multimodal data fusion**, **deep learning**, **time-series modeling**, and **visual explainability**, culminating in robust, interpretable systems for AD progression prediction. The project also explores **adversarial robustness and model resilience**, acknowledging the vulnerability of medical AI systems to misleading or manipulated inputs.

**Core Research Contributions**

**1. Multimodal and Cost-Effective Early Detection**

Our research introduced cost-effective machine learning models that leverage **time-series data from non-invasive and widely available modalities**, including:

* Cognitive scores (MMSE, ADAS, CDRSB, FAQ)
* Demographics
* Medication and comorbidity histories

These models fuse semantically encoded features (e.g., drug data via ATC ontology) with time-sequenced records across 2.5 years. Notably, Random Forest classifiers achieved the highest accuracy for four-class prediction tasks (CN, sMCI, pMCI, AD), even without neuroimaging data.

**2. Multimodal Multitask Deep Learning with Temporal Awareness**

We developed several **hybrid deep learning architectures** combining CNNs and BiLSTMs to model both spatial and temporal dynamics across multimodal time-series data. These models simultaneously:

* Classify patients into diagnostic categories (e.g., CN vs. MCI vs. AD)
* Regress cognitive scores to support fine-grained clinical assessment

This approach outperformed traditional single-modality, single-task models and established new benchmarks on the ADNI dataset.

**3. Explainable and Visual Deep Learning for Medical Trust**

To foster clinical adoption, we proposed a **temporal explainability framework** using guided Grad-CAM over longitudinal 3D MRI scans. Our method tracks how brain regions evolve across time and contributes to prediction decisions, offering voxel-level insights into AD pathology. This time-aware visual XAI module directly supports clinicians in validating model outputs.

**4. Adversarial Robustness in Predictive Frameworks**

Recognizing the susceptibility of deep learning models to adversarial perturbations, our research also evaluated model behavior under **data distortion, limited feature availability, and input manipulation**. We investigated how fused multimodal inputs and multitask learning improve not only predictive performance but also robustness and stability under realistic constraints.

**Project Objectives**

* Develop medically-acceptable deep learning systems for AD progression detection using **multimodal, time-series data**.
* Incorporate **explainability and transparency** into predictive modeling to comply with clinical and regulatory expectations.
* Design **resilient architectures** that maintain high performance in the face of **adversarial conditions and data limitations**.
* Optimize models for deployment in **real-world medical environments**, especially in resource-constrained settings.
* Provide public datasets, visual analytics, and interpretable AI modules to support further research and collaboration.

**Research Impact:** This project bridges the critical gap between cutting-edge AI research and its clinical applicability in the diagnosis of neurodegenerative diseases. By combining **early fusion of multimodal inputs**, **time-aware deep learning**, **adversarial threat analysis**, and **XAI techniques**, the solutions developed by InfoLab at SKKU set a new standard for **transparent, trustworthy, and clinically viable AI systems in healthcare**.

# Project Title: Explainable Dynamic Ensemble Learning with Late Fusion of Multimodal Data for Intelligent Decision Support

**Project Description:** In many high-stakes applications, such as healthcare, finance, and cybersecurity, data is inherently multimodal, originating from diverse sources including images, sensor readings, tabular clinical records, and textual reports. Leveraging such heterogeneous data effectively requires advanced modeling strategies that can handle missing data, modal imbalance, and prediction uncertainty. **Ensemble learning**, particularly **dynamic ensemble selection (DES)**, offers a promising solution by selecting the most competent classifiers for each test instance. Yet, most current DES systems rely solely on **early fusion**, limiting their flexibility, diversity, and performance. Additionally, there is a notable lack of **explainability** tools for dynamic ensemble models, which hinders trust and adoption in sensitive domains.

To address these limitations, **InfoLab at SKKU** has developed a **novel explainable dynamic ensemble learning framework** that utilizes **late fusion of multimodal data**. This project introduces both the conceptual design and practical implementation of new DES techniques, integrating them into an open-source Python library called **Infodeslib**. It further demonstrates superior performance and interpretability across multiple real-world datasets, including critical healthcare and financial applications.

**Key Contributions of the Project**

**1. Infodeslib: Python Library for Dynamic Late Fusion**

* A fully-documented Python package implementing **4 dynamic classifier selection (DCS)** and **7 dynamic ensemble selection (DES)** techniques, adapted for late fusion settings.
* Provides **flexibility in assigning classifiers to different modalities**, enabling better generalization and robustness.
* Implements **explainable AI (XAI)** tools such as:
  + **Case-Based Reasoning (CBR)**: presents similar past cases for comparison.
  + **Classifier Contribution Visuals**: quantifies and visualizes each classifier’s impact on a given prediction.
  + **Local Feature Importance**: uses SHAP values to show which features drove the decision.

**2. Dynamic Late Fusion Framework with Clinical Validation**

* Introduced a new architecture extending **KNORA-U** for late fusion using dynamic selection of classifiers based on region-of-competence strategies.
* Evaluated on the **MIT-GOSSIS in-hospital mortality dataset** (6,600 patients), achieving **90.16% accuracy**, outperforming static and early-fusion dynamic ensembles.
* Results validated across additional datasets such as **ADNI**, **NACC**, **PPMI**, **Credit Card Clients**, and a **neonatal ICU dataset** from the Samarkand Neonatal Center.
* Demonstrated the advantage of late fusion in handling **heterogeneous and missing data**, with better model diversity and generalization.

**Project Objectives**

* Design an **adaptive ensemble learning framework** that supports multimodal decision-level fusion through dynamic selection.
* Provide **transparent and interpretable predictions** using embedded XAI modules for domain experts.
* Evaluate performance across diverse application areas using **benchmark multimodal datasets**.
* Develop an **open-source, user-friendly software library** to encourage reproducibility, collaboration, and real-world deployment.

**Research Impact:**

This research delivers a next-generation solution for intelligent decision support systems by unifying **dynamic selection, multimodal late fusion, and explainable AI**. It equips data scientists and domain experts, especially in healthcare and finance, with powerful tools to make **more accurate, reliable, and understandable predictions** from complex multimodal data. The publicly available **Infodeslib library** serves as a practical resource to accelerate future innovation in dynamic ensemble modeling and trustworthy machine learning.

# Project Title: Explainable and Dynamic Ensemble Models for ICU Mortality and Length-of-Stay Prediction

**Project Description:** Accurate and timely prediction of patient outcomes in Intensive Care Units (ICUs), particularly mortality risk and length of stay (LOS), is crucial for enhancing care quality, reducing hospital burden, and facilitating resource-efficient planning. Traditional scoring systems and even classical machine learning approaches often fall short due to their limited flexibility, poor generalization, or lack of explainability in complex and high-dimensional ICU environments.

To address these limitations, **InfoLab at Sungkyunkwan University (SKKU)** has initiated a research project focused on developing **explainable, dynamic, and high-performing ensemble models** that leverage multivariate time-series data collected during the early hours of ICU admission. This project presents two complementary efforts that push the boundaries of intelligent clinical decision support systems for both adult and neonate ICU patients.

**Core Contributions**

**1. Patient-Specific Stacking Ensemble Model for Adult ICU Mortality Prediction**

* A **novel stacking ensemble framework** is developed using base, each trained on different data modalities determined by medical experts.
* Feature engineering is guided by domain knowledge and includes temporal slices of patient records at 6, 12, and 24 hours post-admission.
* Evaluated on the MIMIC-III dataset with over 10,000 patients, the model achieves **94.4% accuracy** and outperforms traditional scoring systems (e.g., APACHE, SOFA) and popular ML baselines.
* The study emphasizes **data fusion and medical interpretability**, creating a practical pipeline that adapts to the heterogeneous nature of ICU data.

**2. Multilayer Dynamic Ensemble for Neonatal ICU Mortality and LoS Prediction**

* A **two-layer dynamic ensemble system** is proposed for neonates admitted to the NICU, where Layer 1 predicts mortality and Layer 2 estimates length of stay as a regression task.
* Introduces **Dynamic Ensemble Selection (DES)** techniques to adaptively choose the most competent classifiers based on local feature space, addressing the variability and uncertainty in neonatal data.
* Utilizes **explainable AI (XAI) techniques**—including SHAP, decision tree visualization, and rule-based logic—to enhance model transparency and support physician trust.
* Built upon a refined cohort of 3,133 neonates from MIMIC-III, the system integrates a **web-based decision support interface**, making it directly deployable for real-time ICU monitoring.

**Project Objectives**

* Build robust, generalizable models for **early ICU mortality prediction** and **length-of-stay estimation** for both adult and neonatal populations.
* Employ ensemble learning techniques, especially **stacking and dynamic selection**, to combine model diversity with performance.
* Integrate **explainability tools** to foster clinical trust and ensure transparency in model outputs.
* Address real-world issues such as **missing data, time irregularity, and heterogeneous sources** through intelligent preprocessing and domain-driven design.

**Research Impact:** This project demonstrates that combining **medical knowledge with advanced AI techniques** can lead to clinically valuable tools for critical care decision-making. By embedding **interpretable and dynamic ensemble models** into ICU workflow, InfoLab at SKKU aims to:

* Enhance patient safety through earlier risk identification,
* Improve resource allocation by forecasting ICU bed usage,
* Empower medical staff with actionable insights backed by transparent AI decisions.

Together, these studies represent a significant step toward building **trustworthy, data-driven ICU support systems** tailored to the complexities of modern healthcare.