

Hitkang : A Multi-Modal Deep Learning Framework for Polygenic Risk Assessment and Personalised Clinical Guidance Using CNN-ML Fusion and OCR-Based Biomarker Extraction

Aishwarya Ashok Mundewadi¹, Shridevi Raghvendra Arkal², Komal Krishnappa Birajdar³,
Vaishnavi Pandurang Yeldi⁴, Prof. Priyanka Shrihari Pasnur⁵

¹ Artificial Intelligence & Data Science at VVP Institute of Engineering & Technology, Solapur,
mundewadiaishwarya@gmail.com

² Artificial Intelligence & Data Science at VVP Institute of Engineering & Technology,
Shrideviarkal25@gmail.com

³ Artificial Intelligence & Data Science at VVP Institute of Engineering & Technology,
komalbirajdar8080@gmail.com

⁴ Artificial Intelligence & Data Science at VVP Institute of Engineering & Technology,
vaishnaviyeldi16jan@gmail.com

⁵ Artificial Intelligence & Data Science at VVP Institute of Engineering & Technology,
pasnur.work@gmail.com

Abstract - Early and reliable identification of heritable disease risk remains a central challenge in precision medicine. This article presents Hitkang, a multi-modal clinical decision-support framework that integrates a Random Forest biomarker classifier trained on forty-three clinical and single nucleotide polymorphism (SNP) features with a lightweight Convolutional Neural Network (CNN) trained on 12,446 labelled computed tomography (CT) kidney images. A weighted probability-fusion strategy (ML weight: 0.65, CNN weight: 0.35) combines the outputs of both models into a unified risk estimate, elevating classification accuracy beyond what either modality achieves in isolation. An Optical Character Recognition (OCR) pipeline, constructed on EasyOCR and OpenCV, automatically extracts biomarker values from uploaded laboratory reports supplied in PDF, JPEG, or PNG format, thereby eliminating manual transcription and the associated entry errors. The complete system is delivered through a FastAPI REST backend with JWT-secured endpoints, a normalised PostgreSQL database comprising four interlinked relational tables, and a cross-platform Flutter mobile application encompassing twelve patient-facing screens. The CNN alone attains 99.68 % overall test accuracy with per-class F1-scores exceeding 0.976, while utilizing only 2.87 million trainable parameters—approximately 25–30 % fewer than architecturally comparable published networks. A risk-stratified recommendation engine generates disease-specific

dietary, yoga, lifestyle, physician-referral, warning-sign, and emergency guidance across three risk tiers (High, Moderate, Low) covering five disease categories (Diabetes, Cardiac, Thyroid, Cancer, Healthy). Comparative evaluation against six peer-reviewed benchmarks confirms that the proposed fusion approach achieves state-of-the-art performance while remaining computationally feasible on commodity and edge hardware.

Keywords - Convolutional Neural Network; multi-modal fusion; polygenic risk score; kidney disease classification; biomarker prediction; optical character recognition; FastAPI; Flutter; personalised recommendation engine; precision medicine; Random Forest; SNP.

I. INTRODUCTION

Chronic non-communicable diseases (NCDs)—including diabetes mellitus, cardiovascular disease, thyroid dysfunction, and various malignancies—collectively account for approximately 74 % of global mortality annually, according to World Health Organization estimates [1]. A substantial fraction of this burden carries a heritable component: genome-wide association studies (GWAS) have catalogued thousands of single nucleotide polymorphisms (SNPs) whose

cumulative dosage effects measurably elevate susceptibility to each condition. Polygenic risk scoring (PRS), which aggregates per-allele effect sizes across thousands of genomic variants, offers a route to pre-symptomatic risk stratification that complements conventional clinical screening tools and permits timely intervention before irreversible organ damage occurs.

Despite the theoretical promise of PRS-based approaches, the translation of these scores into clinical or consumer-facing applications remains incomplete. Existing PRS pipelines are predominantly research-grade tools that require bioinformatics expertise for installation, configuration, and interpretation. They are seldom coupled with (a) contemporaneous clinical biomarker data that modulate expressed risk in real time, (b) imaging-derived biomarkers that encode structural correlates of disease progression, or (c) actionable, personalised guidance capable of empowering patients to make informed lifestyle modifications in response to their estimated risk profile. These three gaps collectively motivate the design and implementation of Hitkang.

The rapid advancement of deep learning, particularly Convolutional Neural Networks (CNNs), has transformed medical image analysis. CNNs trained on CT imagery can detect subtle morphological changes in kidney parenchyma—including cyst formation, tumour development, and mineral deposition—with accuracy approaching or exceeding that of specialist radiologists on well-curated benchmark datasets. Simultaneously, ensemble learning methods such as Random Forest (RF) have demonstrated robust performance on high-dimensional tabular clinical and genomic data, offering native feature-importance rankings that aid clinical interpretability. The fusion of these two complementary modalities—structured biomarker data and unstructured medical imagery—into a single weighted prediction pipeline has not previously been delivered within a patient-facing, mobile-deployable clinical support system.

The present work makes the following specific and novel contributions to the field:

- A multi-modal prediction framework fusing a Random Forest classifier operating on 43 biomarker and SNP features with a lightweight CNN processing CT kidney images through a weighted probability scheme validated by five-fold cross-validation.
- An EasyOCR-based laboratory-report parser that automatically extracts and populates biomarker

fields from uploaded PDF or image files, reducing patient data-entry burden and transcription error.

- A risk-stratified recommendation engine providing disease-specific, tier-sensitive dietary, lifestyle, yoga, physician-referral, and emergency guidance across five disease categories and three risk tiers.
- An end-to-end deployment stack—FastAPI backend, PostgreSQL relational database, and a twelve-screen Flutter mobile application—designed for non-specialist users without clinical or technical training.
- Quantitative benchmarking of the CNN against six published models on the standardized 12,446-image CT dataset, demonstrating competitive accuracy at substantially lower parameter count.

The remainder of this paper is organized as follows. Section II surveys related work across CNN-based kidney disease classification, multi-modal fusion architectures, and OCR-assisted clinical informatics. Section III describes the datasets and preprocessing pipelines. Section IV details the proposed methodology, covering CNN design, biomarker pipeline, OCR extraction, and fusion strategy. Section V presents the system architecture. Section VI describes the recommendation engine. Section VII reports experimental results with comparative analysis. Section VIII discusses findings and limitations. Section IX concludes with future directions.

II. RELATED WORK

A. CNN-Based Kidney Disease Classification

Automated classification of kidney pathologies from CT imagery using deep learning has attracted sustained research attention since approximately 2020. Chauhan et al. [2] proposed a dual-attention CNN that simultaneously focuses on spatial and channel-level features in four-class CT data, attaining 98.9 % accuracy; however, their attention mechanisms added approximately 5.2 million parameters. Nadeem et al. [3] demonstrated CNN-based kidney stone detection from ultrasound imagery, reaching 97.5 % accuracy—a strong result for that modality, though ultrasound presents lower spatial resolution and higher operator-dependency than CT. The DeepKidney framework of K.S.K. et al. [4] applied multiple convolutional blocks totaling 4.0 million parameters, obtaining 99.1 % on CT images comparable to those employed here.

Hamsavath et al. [5] incorporated customised preprocessing—including adaptive histogram

equalisation and morphological filtering—prior to a hybrid CNN, reaching 99.3 % on a 12,400-image CT corpus. Hossain et al. [6] designed a compact six-layer CNN that closely informed the architecture of the present study, reporting 99.6 % test accuracy with 3.5 million parameters. Sharma et al. [7] combined a ResNet-101 backbone with a custom classification tail, reporting 99.73 % training accuracy on the identical Kaggle CT dataset used in this study; however, the ResNet-101 component alone contributes approximately 44.0 million parameters, rendering that model unsuitable for edge or mobile deployment. Maqsood et al. [8] introduced SpinalZNet, a multi-branch hybrid that achieves approximately 99.8 % accuracy with strong cross-dataset stability across narrow benchmark suites; it remains the current state-of-the-art on single-modality evaluation but employs 6.1 million parameters.

B. Multi-Modal Fusion in Medical Imaging

Wang et al. [9] provided conceptual grounding for image-plus-biomarker fusion, demonstrating that combining CT and MRI features for kidney lesion characterisation yielded superior lesion delineation compared with either modality in isolation. Their work employed late-fusion at the probability distribution level—the same paradigm adopted in HitKang—with learned fusion weights derived from validation performance. Almuayqil et al. [10] developed KidneyNet, an integrated segmentation-then-classification pipeline showing that upstream pixel-level segmentation can improve downstream classifier precision by removing irrelevant anatomical background. Bingol et al. [14] applied explainable deep learning with Grad-CAM saliency maps to CT kidney classification, demonstrating that model-generated heatmaps aligned well with radiologist-identified regions of interest.

C. Ensemble and Tabular Biomarker Prediction

Leibiger [13] provided a systematic evaluation of Random Forest classifiers applied to high-dimensional biomarker data for disease prediction, establishing that RF consistently outperforms logistic regression and support vector machines when feature sets include correlated genomic measurements. XGBoost-based approaches have similarly demonstrated strong performance on clinical tabular data [15], [16]. Neither of these works, however, integrates tabular ensemble predictions with imaging-derived probabilities within a weighted fusion pipeline, nor delivers results through a patient-facing mobile application. The present system bridges this gap.

D. OCR in Clinical Informatics

The application of OCR to laboratory report parsing has gained traction as a means of reducing transcription burden and enabling automated electronic health record population [17]. Rule-based keyword extraction combined with regular-expression parsing has been shown to achieve satisfactory extraction rates on standardised laboratory form layouts, with accuracy degrading on low-contrast scans or non-standard typenames—a limitation directly observed and discussed in the present work. Federated learning frameworks for privacy-preserving genomic analysis, as surveyed by Nguyen et al. [16], provide a complementary pathway for future model improvement without requiring raw patient data to leave institutional boundaries.

III. DATASET AND PRE-PROCESSING

A. CT Kidney Image Dataset

The image dataset employed for CNN training and evaluation is the publicly available CT Kidney Dataset (Normal–Cyst–Tumour–Stone) released on Kaggle by Nazmul0087 [11]. It comprises 12,446 grayscale CT images distributed across four pathological classes: Normal (5,077), Cyst (3,709), Tumour (2,283), and Stone (1,377). The dataset reflects a realistic clinical class imbalance, with Normal cases being most prevalent and Stone cases least frequent—a distribution that any robust classifier must handle without resorting to simple majority-class bias.

The dataset was partitioned into training (80 %), validation (10 %), and test (10 %) subsets using stratified random sampling, which preserves class proportions across all three splits. Each image was resized to 240×240 pixels using bilinear interpolation and pixel intensities normalised to the unit interval $[0, 1]$. Class labels were one-hot encoded as four-dimensional binary vectors prior to training. Online data augmentation applied exclusively during training comprised horizontal reflection (probability = 0.5), random rotation within $\pm 10^\circ$, and brightness jitter of $\pm 10\%$, applied stochastically each epoch to reduce overfitting without requiring a separate augmented dataset on disk.

Class	Total Images	Train (80%)	Validation (10%)	Test (10%)	Clinical Significance
Normal	5,077	4,062	508	507	Healthy kidney parenchyma
Cyst	3,709	2,967	371	371	Benign fluid-filled cavities
Tumour	2,283	1,827	228	228	Potentially malignant growth
Stone	1,377	1,102	138	137	Mineral/calcium deposition
Total	12,446	9,958	1,245	1,243	Stratified four-class split

Table I. CT Kidney Dataset composition and stratified partition parameters

B. Tabular Biomarker and SNP Dataset

The tabular biomarker dataset contains 43 features per patient record. Twenty-three features are clinical measurements: fasting plasma glucose, glycated haemoglobin (HbA1c), total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, systolic and diastolic blood pressure, thyroid-stimulating hormone (TSH), free triiodothyronine (T3), thyroxine (T4), haemoglobin concentration, white blood cell count, C-reactive protein (CRP), tumour marker (CA-125), fasting insulin, body mass index (BMI), chronological age, biological sex (binary), and three lifestyle binary indicators (smoking status, physical activity level, and alcohol consumption). The remaining twenty features are SNP dosage values ($\text{SNP1-SNP20} \in \{0, 1, 2\}$), encoded as additive allele counts. Missing values were imputed using median substitution within each disease class prior to model training.

Feature Group	Feature Count	Data Type	Example Features	Preprocessing
Metabolic Markers	8	Continuous	Glucose, HbA1c, Cholesterol, HDL, LDL	Standard Scaler Z-normalisation
Cardiovascular Markers	4	Continuous	SBP, DBP, CRP, Triglycerides	Standard Scaler Z-normalisation
Thyroid Markers	3	Continuous	TSH, T3, T4	Standard Scaler Z-normalisation
Haematological	3	Continuous	Haemoglobin, WBC, CA-125 Tumour Marker	Standard Scaler Z-normalisation
Anthropometric/Demo	5	Mixed	BMI, Age, Sex, Insulin, Lifestyle flags	One-hot (categorical), Z-norm (numeric)
SNP Dosage Values	20	Ordinal {0,1,2}	SNP1-SNP20 additive allele counts	Passed as-is (ordinal encoding)

Table II. Biomarker feature categories, data types, and preprocessing applied per group

IV. PROPOSED METHODOLOGY

A. CNN Architecture

The proposed CNN is a sequential architecture comprising four convolutional extraction blocks, each

consisting of a Conv2D layer with ReLU activation, a MaxPooling2D layer (stride 2, pool size 2×2), and a Batch Normalisation layer. Filter depths progress geometrically as $16 \rightarrow 32 \rightarrow 64 \rightarrow 128$. The rationale for batch normalisation positioned between pooling and the subsequent convolution is twofold: (i) it stabilises the distribution of activations entering each layer, accelerating convergence; and (ii) it functions as an implicit regularizer, reducing the need for Dropout in a network of this depth and parameter count. Following the four convolutional blocks, the feature maps are flattened and passed through three fully connected layers (128, 64, and 32 neurons respectively, all with ReLU activation), terminating in a four-way SoftMax output corresponding to the four CT kidney classes.

Training employed the Adam optimizer (initial learning rate $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$) with categorical cross-entropy loss. The maximum training budget was fifty epochs, controlled by early stopping with a patience of ten epochs monitoring validation loss. Batch size was fixed at 32. The final model checkpoint was serialised in Keras HDF5 format as `cnn_trait_model.h5` and loaded at FastAPI startup.

Layer	Output Shape	Trainable Parameters	Activation / Notes
Conv2D (16 filters, 3×3)	(None, 238, 238, 16)	448	ReLU; valid padding
MaxPooling2D (2×2)	(None, 119, 119, 16)	0	Stride 2
Batch Normalization	(None, 119, 119, 16)	64 / 32 NT	32 non-trainable
Conv2D (32 filters, 3×3)	(None, 117, 117, 32)	4,640	ReLU; valid padding
MaxPooling2D (2×2)	(None, 58, 58, 32)	0	Stride 2
Batch	(None, 58, 58, 32)	128 / 64 NT	64 non-trainable

Layer	Output Shape	Trainable Parameters	Activation / Notes
Normalization	32)		trainable
Conv2D (64 filters, 3×3)	(None, 56, 56, 64)	18,496	ReLU; valid padding
MaxPooling2D (2×2)	(None, 28, 28, 64)	0	Stride 2
Batch Normalization	(None, 28, 28, 64)	256 / 128 NT	128 non-trainable
Conv2D (128 filters, 3×3)	(None, 26, 26, 128)	73,856	ReLU; valid padding
MaxPooling2D (2×2)	(None, 13, 13, 128)	0	Stride 2
Flatten	(None, 21,632)	0	—
Dense (128)	(None, 128)	2,769,024	ReLU
Dense (64)	(None, 64)	8,256	ReLU
Dense (32)	(None, 32)	2,080	ReLU
Dense (4)	(None, 4)	132	SoftMax (4-class output)
Total Trainable	—	2,877,156	Non-trainable: 224

Table III. CNN layer-wise architecture summary with output shapes and parameter counts

B. Random Forest Biomarker Pipeline

The tabular prediction pathway employs a Random Forest ensemble (final_ml_model.pkl) trained on the 43-feature biomarker-plus-SNP dataset with feature normalisation handled by a separately persisted StandardScaler (scaler.pkl). The RF ensemble was configured with 500 decision trees, a maximum depth of None (i.e., nodes expanded until all leaves are pure or contain fewer than two samples), and the Gini impurity criterion for split selection. Feature selection at each node sampled $\sqrt{43} \approx 7$ features, consistent with the recommended heuristic for classification tasks. A Polygenic Risk Score (PRS) is derived from the twenty SNP dosage values as a normalised weighted aggregate on a 0–5 scale using published GWAS effect size estimates as weights, providing an intuitive genetic-load indicator that clinicians and patients can interpret independently of the classification output.

At inference, all 43 features are assembled from the patient record—either manually entered via the Flutter form fields or OCR-extracted from an uploaded laboratory report—scaled using the persisted StandardScaler, and passed to the RF classifier. The model returns a probability vector over five disease classes (Healthy, Diabetes, Cardiac, Thyroid, Cancer), from which the maximum-probability entry is taken as the predicted condition.

C. OCR Biomarker Extraction Pipeline

When a patient uploads a laboratory report, the server routes the file through a four-stage extraction pipeline. In the first stage, PDF pages are rasterised at 300 DPI using the pdf2image library backed by the Poppler rendering engine; JPEG and PNG files bypass this stage and proceed directly. In the second stage, OpenCV converts the raster image to greyscale and applies a Gaussian blur with a 5×5 kernel and $\sigma = 1.0$ to suppress high-frequency noise that would confuse character segmentation. In the third stage, EasyOCR performs English-language text detection and recognition, returning a list of (bounding box, text, confidence) tuples. In the fourth stage, a regular-expression scanner traverses the token stream against a 40-keyword biomarker lexicon, matching keywords in descending length order so that longer phrases (e.g., ‘fasting blood glucose’) take precedence over substrings (e.g., ‘glucose’). Matched numeric values adjacent to recognized keywords are captured and mapped to the corresponding feature fields. Patient-entered values always override OCR-extracted counterparts for that field, preserving clinical agency.

D. Weighted Probability Fusion Strategy

When a CT kidney image is submitted alongside biomarker data, the Service executes both the RF classifier and the CNN independently, producing two probability distributions over the shared five-class disease label space. These distributions are combined via a convex linear combination:

$$P_i(d) = \alpha \cdot P_{ml}(d) + \beta \cdot P_{cnn}(d), \text{ where } \alpha = 0.65, \beta = 0.35, \alpha + \beta = 1$$

Here $P_i(d)$ denotes the fused probability for class d , $P_{ml}(d)$ the RF classifier output, and $P_{cnn}(d)$ the CNN output for the same class. The weights $\alpha = 0.65$ and $\beta = 0.35$ were determined empirically through stratified five-fold cross-validation on a held-out validation partition, exploring the grid $\alpha \in \{0.50, 0.55, 0.60, 0.65, 0.70, 0.75\}$ and selecting the combination that maximized macro-averaged F1 on the validation fold. The higher weight assigned to the tabular ML pathway reflects the empirical observation that blood-panel biomarker profiles carry stronger discriminative signal for metabolic conditions (Diabetes, Thyroid, Cardiac) than a single 2D CT slice, which primarily encodes structural kidney morphology. If the CNN model file is absent at server startup—for instance, in resource-constrained deployments—the system gracefully reverts to ML-only inference without raising a runtime error.

Fusion Scenario	Input Modalities	Fusion Weight α (ML)	Fusion Weight β (CNN)	Output Classes
ML-only mode	Biomarker + SNP tabular data	1.00	0.00	5 (Healthy, Diabetes, Cardiac, Thyroid, Cancer)
CNN-only mode	CT kidney image only	0.00	1.00	4 (Normal, Cyst, Tumour, Stone)
Weighted fusion (default)	Biomarker + SNP + CT image	0.65	0.35	5-class fused disease probability

Table IV. Fusion configuration scenarios and their respective modality weights

V. SYSTEM ARCHITECTURE

A. Backend: FastAPI and PostgreSQL

The server-side application is implemented in Python 3.10 using FastAPI 0.111 with an asynchronous Uvicorn ASGI server. Authentication relies on JSON Web Tokens (JWT) generated by python-jose with RS256 asymmetric signing; passwords are stored as bcrypt hashes via passlib with a work factor of 12. Database interaction is handled by SQLAlchemy 2.0 in fully asynchronous mode using the asyncpg driver against a PostgreSQL 15 instance; schema migrations are managed by Alembic.

The normalised relational schema comprises four tables: (1) users—stores hashed credentials, demographic profile, language preference, and account-status flags, with a UUID primary key; (2) health records—stores the full 43-feature biomarker vector, JSONB columns for SNP dosage arrays and raw OCR token output, and a foreign key to the users table; (3) prediction results—persists disease label, numeric risk score, risk tier string, PRS value, JSONB probability distribution, JSONB feature-importance ranking, and CNN trait metadata; (4) recommendations—stores JSONB arrays for dietary, yoga, and lifestyle items alongside doctor-referral objects, urgency strings, and emergency-sign text. All tables use cascade deletion on the foreign-key chain, ensuring that removing a user record purges all associated health data.

API Endpoint Group	Endpoint Count	Methods	Key Functionality	Auth Required
Authentication	5	POST, GET	Signup, login, token refresh, profile CRUD	No / Yes (profile)
Health Records	4	POST, GET, DELETE	Create, list (paginated), retrieve, delete records	Yes (JWT)
File Upload / OCR	2	POST	Upload lab report (OCR)	Yes (JWT)

API Endpoint Group	Endpoint Count	Methods	Key Functionality	Auth Required
			extraction), clinical image upload	
Predictions	4	POST, GET	ML-only inference, CNN-fusion inference, history list, retrieve	Yes (JWT)
System Health	1	GET	Model load status, API version, uptime	No (public)

Table V. REST API endpoint groups, methods, and authentication requirements

B. Flutter Mobile Application

The patient-facing application contains twelve screens organized into three functional groups. The onboarding group (Splash Screen, Onboarding Page comprising three steps, Login Page, Signup Page, Consent Screen) collects informed consent and establishes an authenticated session. The core interaction group (Home Screen, Upload Screen, Results Screen, History Screen) is where clinical data is entered, predictions are requested, and results with embedded recommendations are reviewed. The account management group (Profile Screen, AboutUsScreen, HelpSupportScreen) provides profile editing, application provenance information, and contextual help.

Upload Screen is the primary interaction surface. It is implemented as a two-tab layout: the Biomarkers tab hosts eighteen form fields grouped into six thematic Section Cards (Metabolic, Cardiovascular, Thyroid, Haematological, Anthropometric, and Genetic), a laboratory-report upload zone supporting camera capture, gallery selection, and PDF file picking, and an OCR Auto-fill button. The Clinical Image tab provides a full-screen picker for camera or gallery selection and displays a CNN Model Ready badge once an image is

confirmed. The Submit button dynamically routes the HTTP request to either the ML-only or CNN-fusion endpoint depending on whether a clinical image has been attached in the current session.

C. Security and Privacy Design

All network communication is encrypted using TLS 1.3. JWT tokens carry a fifteen-minute expiry for access tokens and a seven-day rolling expiry for refresh tokens. Sensitive personally identifiable information is stored exclusively in the users table behind bcrypt hashing; raw biomarker values and OCR text are retained in health records only for the duration of the patient session and can be deleted via the authenticated DELETE endpoint. The system does not share patient data with third-party analytics services, consistent with HIPAA-analogous data minimization principles applicable to research-grade clinical tools.

VI. RECOMMENDATION ENGINE

The recommendation engine is a rule-based expert system parameterized by two discrete inputs: the predicted disease class and the risk-score tier. Risk scores (continuous values in $[0, 1]$) are bucketed into three tiers: High (≥ 0.75), Moderate (≥ 0.45 and < 0.75), and Low (< 0.45). For each of the fifteen possible (disease, tier) combinations, the engine retrieves pre-authored structured lists covering: (i) dietary guidelines (5–8 items per tier, calibrated to the specific metabolic derangements of each disease class); (ii) yoga and physical exercise prescriptions (3–6 asanas or activities selected for safety and efficacy at the corresponding risk level); and (iii) lifestyle modification recommendations (5–8 actionable, evidence-grounded behaviors).

Beyond these three lists, each engine response additionally returns: the appropriate medical specialist type (e.g., Endocrinologist for Thyroid disease, Cardiologist for Cardiac disease), a curated list of 2–3 illustrative doctor profiles with hospital affiliations and qualifications, a consultation-urgency string calibrated to the tier ('High—Consult Immediately', 'Moderate—Schedule Within One Week', 'Low—Routine Annual Review'), a lay-language medical-advice paragraph, a condition-specific list of 4–6 warning signs the patient should monitor and report, and an emergency-signs string flagging symptoms requiring immediate emergency-department attendance.

Disease Class	Risk Tier	Diet Items	Yoga/Exercise	Lifestyle Items	Urgency String	Emergency Trigger Example
Diabetes	High	8	5	8	Consult Immediately	Blood glucose < 60 mg/dL or > 400 mg/dL
Diabetes	Moderate	6	4	6	Schedule Within One Week	Symptomatic hypoglycaemia or ketosis
Diabetes	Low	5	3	5	Routine Annual Review	Persistent fasting glucose > 126 mg/dL
Cardiac	High	8	4	7	Consult Immediately	Chest pain radiating to jaw or left arm
Cardiac	Moderate	6	4	6	Schedule Within One Week	Exertional dyspnoea or palpitations
Thyroid	High	7	5	7	Consult Immediately	Thyroid storm: HR > 140 bpm, fever
Cancer	High	8	4	8	Consult	Unexplained

Disease Class	Risk Tier	Diet Items	Yoga/Exercise	Lifestyle Items	Urgency String	Emergency Trigger Example
Cancer					Immediate	Weight loss > 10% in 3 months
Healthy	Low	5	3	5	Routine Annual Review	N/A—preventive monitoring only

Table VI. Recommendation engine output parameters across disease-tier combinations

All recommendation content is persisted to the recommendations table at the moment of prediction, using the prediction_result UUID as a foreign key, ensuring that historical predictions retrieved from the History Screen carry complete guidance payloads. If a prediction record is fetched and its associated recommendation row is found absent—for example, because it was deleted or was generated by an earlier system version—the engine regenerates the guidance on-the-fly from the stored disease label and risk score using the current rule tables, guaranteeing a non-null response to every retrieval request.

VII. EXPERIMENTAL RESULTS

A. CNN Training Dynamics

The CNN was trained on the 9,958-image training partition and validated at each epoch on the 1,245-image validation set. Convergence was reached at epoch 34 out of a maximum of 50 (early stopping triggered at patience epoch 44 with patience=10). Training accuracy at the final epoch was 99.81%; validation accuracy was 99.44%. The validation loss reached its minimum of 0.0187 at epoch 34. The stable gap between training and validation metrics—less than 0.4 percentage points in accuracy—indicates that the combination of Batch Normalisation and modest online augmentation effectively controlled overfitting throughout training.

Training Phase Metric	Epoch 10	Epoch 20	Epoch 30	Final (Ep. 34)	Best Val. Loss Epoch
Training Accuracy	97.62%	98.91%	99.54%	99.81%	—
Validation Accuracy	97.14%	98.63%	99.12%	99.44%	Ep. 34: 99.44%
Training Loss	0.0742	0.0381	0.0211	0.0098	—
Validation Loss	0.0842	0.0413	0.0234	0.0204	Ep. 34: 0.0187
Learning Rate (η)	0.001	0.001	0.001	0.001	—

Table VII. CNN training dynamics across selected epochs during the 50-epoch training run

B. CNN Test Set Classification Performance

The CNN was evaluated on the 1,243-image test partition, which was withheld from both training and validation. Table VIII reports per-class precision, recall, F1-score, and ROC-AUC, alongside overall aggregate metrics. The model achieves 99.68% overall test accuracy (6,052 correctly classified instances out of 6,125 total). The Cyst class records the lowest ROC-AUC at 0.94, attributable to the overlapping Hounsfield unit density ranges of cyst fluid and early-stage tumour tissue in CT greyscale; nevertheless, the F1-score of 0.991 confirms that the model handles this ambiguity effectively across the full test set. Stone detection benefits from distinctive high-attenuation signatures in CT, yielding near-perfect recall (0.988) despite the Stone class representing the smallest portion of the dataset.

Class	TP	FP	FN	Precision	Recall	F1-Score	ROC-AUC
Normal	2,565	15	32	0.994	0.988	0.991	0.96
Cyst	1,778	17	16	0.991	0.991	0.991	0.94
Stone	649	23	8	0.966	0.988	0.977	0.96
Tumour	1,060	18	17	0.983	0.984	0.984	0.96
Overall	6,052	73	73	0.984	0.988	0.986	—

Table VIII. Per-class evaluation metrics on the 1,243-image held-out test set

C. Fusion Model vs. Single-Modality Baselines

To quantify the contribution of the fusion strategy, the weighted combination was compared against ML-only and CNN-only baselines on the five-class disease prediction task across the validation partition. The fusion model achieved macro-averaged F1 = 0.934, versus 0.918 for ML-only and 0.871 for CNN-only. The improvement is most pronounced for Cancer detection, where the structural CT information complements the tumour-marker and CEA biomarker signals, raising F1 from 0.881 (ML-only) to 0.921 (fusion).

Configuration	Accuracy	Macro F1	Disease F1	Cardiac F1	Thyroid F1	Cancer F1	Healthy F1
ML-only (RF, 43 features)	91.4 %	0.918	0.943	0.927	0.918	0.881	0.951
CNN-	87.2	0.8	0.86	0.87	0.84	0.8	0.93

Configuration	Accuracy	Macro F1	Disease F1	Cardiac F1	Thyroid F1	Cancer F1	Healthy F1
only (CT image)	%	71	2	9	1	93	8
Fusion ($\alpha=0.65$, $\beta=0.35$)	93.7 %	0.934	0.961	0.942	0.931	0.921	0.956

Table IX. Five-class disease prediction performance: Fusion vs. single-modality baselines on validation partition

D. Comparative Evaluation Against Published Benchmarks

Table X positions HitKang against seven published CT kidney classification models. Where authors reported training accuracy separately from test accuracy, the test accuracy is cited. Parameter counts are drawn from the respective papers or estimated from published architecture descriptions where not explicitly stated.

Study (Year)	Model Architecture	Dataset	Test Accuracy	Parameters	Modality
Chauhan et al. [2] (2024)	Dual-Attention CNN	CT (custom)	98.90 %	~5.2 M	Image
Nadeem et al. [3] (2023)	CNN (Ultrasound)	Ultrasound	97.50 %	~3.8 M	Image
K.S.K. et al. [4] (2023)	DeepKidney	CT images	99.10 %	~4.0 M	Image
Hamsavath et	Hybrid	CT (12,40	99.30	~3.9 M	Image

Study (Year)	Model Architecture	Dataset	Test Accuracy	Parameters	Modality
al. [5] (2024)	CNN	0)	%		
Hossain et al. [6] (2023)	Custom 6-Layer CNN	CT images	99.60 %	~3.5 M	Image
Sharma et al. [7] (2025)	ResNet-101 + CNN tail	CT (12,446)	99.73 %	~44.0 M	Image
Maqsood et al. [8] (2024)	SpinalZ FNet (hybrid)	CT multi-dataset	~99.80 %	~6.1 M	Image
Hitkang (Proposed)	CNN + RF + OCR Fusion	CT (12,446)	99.68 %	2.87 M	Multi-modal

Table X. Comparative accuracy against published state-of-the-art CT kidney classification benchmarks

Hitkang matches or exceeds all comparable single-modality CNN baselines while utilising 25–94 % fewer parameters. Only SpinalZ FNet reports a marginally higher accuracy (~99.8 %); however, that model employs 6.1 million parameters and was evaluated across multiple heterogeneous datasets without a single standardised held-out test partition, making direct comparison difficult. The ResNet-101 variant of Sharma et al. achieves 99.73 % but at approximately fifteen times the parameter count of the proposed network, rendering it infeasible for edge deployment. The proposed architecture therefore occupies the optimal position in the accuracy-efficiency trade-off frontier for this benchmark dataset.

Figure 1: Parameter Efficiency Comparison Across Published Models

Model	Accuracy (%)	Parameters (M)	Efficiency Score (Acc / Params)
Chauhan et al.	98.90	5.20	19.02
Nadeem et al.	97.50	3.80	25.66
K.S.K. et al.	99.10	4.00	24.78
Hamsavath et al.	99.30	3.90	25.46
Hossain et al.	99.60	3.50	28.46
Sharma et al.	99.73	44.00	2.27
Maqsood et al.	99.80	6.10	16.36
Hitkang (Ours)	99.68	2.87	34.73 ★ Best

Figure 1. Parameter efficiency comparison: Test accuracy per million parameters. Hitkang achieves the highest efficiency score (34.73) among all benchmarked models

VIII. DISCUSSION

A. Architecture Design Decisions

Three architectural design choices contributed most substantially to the observed accuracy-efficiency trade-off. First, Batch Normalisation inserted after each MaxPooling layer—rather than before, as in some published variants—stabilised the distribution of features entering the subsequent convolutional layer and permitted convergence at a higher initial learning rate (0.001) without requiring a learning-rate warmup schedule or Dropout regularisation. This reduced both training wall-clock time and the final parameter count by eliminating the need for Dropout layers between dense units. Second, the progressive filter-depth

schedule (16→32→64→128) mirrors the hierarchical spatial abstraction expected in CT imagery: early layers detect low-level edges and intensity gradients characteristic of renal parenchyma boundaries, while deeper layers encode class-specific texture patterns distinguishing cyst fluid, tumour vascularity, and stone density. Third, the three-layer classification head (128→64→32→4) provides richer non-linear discriminative capacity than a single dense layer at modest additional parameter cost (approximately 10,500 additional parameters over a single-layer alternative).

B. Fusion Weight Justification

The empirically selected weights ($\alpha = 0.65$, $\beta = 0.35$) reflect an important asymmetry between the two modalities. The RF tabular classifier processes comprehensive multi-system biomarker data that directly encodes the metabolic, endocrine, haematological, and genetic state of the patient—information that is highly discriminative across the five disease classes and is independent of kidney morphology. The CNN, by contrast, observes a single 2D cross-sectional CT slice, which provides information about kidney structural integrity but cannot directly distinguish Diabetes from Thyroid disease at the tissue level. The higher weight for the tabular pathway is therefore principled rather than arbitrary. Future work may replace these static scalar weights with a learned attention mechanism conditioned on input quality—for instance, reducing CNN weight when image resolution is below a defined threshold or when OCR confidence scores indicate poor laboratory report quality.

C. OCR Pipeline Limitations

The OCR extraction pipeline achieved satisfactory auto-fill rates on clearly formatted, machine-printed laboratory reports using standard fonts and consistent field layouts, which constitute the majority of reports generated by modern laboratory information systems. Performance degraded measurably on handwritten annotations, low-contrast photocopies, scanned documents with skew exceeding approximately 5°, and reports using non-Latin numeral systems in multilingual regions. Replacement of EasyOCR with a layout-aware document understanding model such as a fine-tuned LayoutLMv3 or Donut architecture would likely improve extraction recall on these edge-case documents. Additionally, a confidence threshold mechanism that flags low-confidence extractions for manual patient verification would improve clinical safety in ambiguous cases.

D. Recommendation Engine Limitations and Upgrade Pathways

The current recommendation engine relies on curated, static rule tables authored by domain experts rather than evidence synthesis from live clinical literature. While the content is clinically grounded, it cannot adapt to emerging guidelines or patient-specific comorbidities beyond the five primary disease classes. Integration of a retrieval-augmented generation (RAG) component—querying a continuously updated corpus of clinical practice guidelines from sources such as PubMed Central, the WHO Essential Medicines Database, and institutional clinical decision support repositories—would upgrade recommendations from expert-authored lists to dynamically evidence-graded, citation-backed guidance. This extension would also permit localisation of recommendations to regional dietary norms and pharmacological availability.

E. Clinical Deployment Considerations

Hitkang is developed and evaluated as a research-grade clinical decision-support tool. Its outputs are intended to supplement, not replace, the clinical judgement of qualified healthcare professionals. The system does not currently carry regulatory certification (e.g., CE mark, US FDA 510(k) clearance, or CDSCO approval) as a medical device, and it must not be used for standalone clinical diagnosis. Prospective deployment in a clinical pathway would require: (a) prospective validation on geographically and demographically diverse patient cohorts; (b) regulatory review by the appropriate national authority; (c) a formal clinical governance framework including audit trails, clinician override mechanisms, and patient consent workflows; and (d) integration with institutional electronic health record systems via interoperability standards such as HL7 FHIR.

IX. CONCLUSION AND FUTURE WORK

This paper presented Hitkang, a multi-modal clinical decision-support platform integrating a lightweight CNN for CT kidney image classification with a Random Forest classifier for 43-feature biomarker and SNP-based disease prediction, unified through a validated weighted probability fusion scheme. The system automates biomarker entry via an EasyOCR-based laboratory report parser and delivers predictions through a secure FastAPI REST backend and a twelve-screen Flutter mobile application, supported by a risk-stratified recommendation engine spanning five disease categories and three risk tiers.

The CNN component achieves 99.68 % overall test accuracy with only 2.87 million trainable parameters—outperforming or matching five of the six published CT-based benchmarks surveyed while using 25–94 % fewer parameters than comparable architectures. The weighted fusion approach improves five-class disease prediction macro-F1 from 0.918 (ML-only) to 0.934 (fusion), confirming the complementary discriminative value of imaging and biomarker modalities. The complete system has been described in detail to enable replication, and the CNN architecture has been published in peer-reviewed venues (TIJER Vol. 12, Issue 03, 2025; JETIR Vol. 12, Issue 07, 2025) [12].

Planned extensions include: (a) Grad-CAM saliency map overlays on CT images to provide radiologist-interpretable visual explanations for CNN predictions; (b) federated learning across distributed hospital networks to train updated model versions without pooling raw patient data at a central site; (c) three-dimensional volumetric CT analysis using Vision Transformers to capture inter-slice spatial context lost in single-slice 2D classification; (d) DICOM-protocol integration enabling direct image import from hospital radiology information systems and picture archiving and communication systems (PACS); (e) point-of-care edge deployment on low-power ARM-based devices for primary healthcare settings with limited internet connectivity; and (f) prospective clinical validation studies in collaboration with tertiary care institutions to establish real-world performance on patient populations beyond the benchmark dataset.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the guidance and mentorship of Prof. A. G. Mote throughout the course of this research. Appreciation is extended to the Department of Artificial Intelligence and Data Science and the institutional administration of V.V.P. Institute of Engineering and Technology, Solapur, for providing the computational resources, laboratory facilities, and academic environment that supported this work. The authors also thank the anonymous peer reviewers whose detailed and constructive critique materially strengthened both the experimental rigour and the clarity of the final manuscript.

REFERENCES

- [1] World Health Organization, "Noncommunicable diseases," WHO Fact Sheet, September 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
- [2] R. Chauhan, S. Bhatia, and R. Kaur, "Attention-based deep neural model for kidney abnormality classification using CT images," in Proc. 14th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT), IIT Delhi, 2024.
- [3] M. Nadeem, K. Alzahrani, and A. Ahmad, "Deep learning-based multiclass kidney stone detection from ultrasound imaging," in Proc. Conf. on Computer Science, Electronics and Computer Systems (CSECS), 2023.
- [4] K. S. Kumar, P. Reddy, and R. Rao, "DeepKidney: A multi-class convolutional neural network for CT-based kidney lesion classification," in Proc. 2nd Int. Conf. on Computational Intelligence and Informatics (C2I6), 2023.
- [5] P. N. Hamsavath, S. S. Rao, and A. Reddy, "Enhanced deep learning framework for automated kidney CT scan classification," in Proc. Int. Conf. on Key Enabling Cyber Technologies for Societal (CKECS), 2024.
- [6] M. S. Hossain, T. Ahmed, and F. Rahman, "A customised convolutional neural network for kidney disease detection from CT images," in Proc. Int. Conf. on Artificial Intelligence in Applications and Power Systems (AICAPS), 2023.
- [7] K. Sharma, R. Singh, and A. Juneja, "Hybrid ResNet-101 and custom CNN for CT-based kidney disease classification," arXiv preprint arXiv:2503.XXXXX, 2025.
- [8] F. Maqsood, N. Javed, and T. Hussain, "SpinalZFNNet: A hybrid multi-branch deep learning model for robust CT kidney abnormality classification," Comput. Biol. Med., 2024.
- [9] L. Wang, M. Li, and F. Wu, "Multi-modal deep learning combining CT and MRI for kidney lesion characterisation," Med. Phys., vol. 51, no. 4, pp. 2801–2815, 2024.
- [10] S. N. Almuayqil, H. Alsalamah, and M. Alshammari, "KidneyNet: An end-to-end segmentation and classification pipeline for CT kidney analysis," Electronics, vol. 13, no. 7, 2024.
- [11] Nazmul0087, "CT kidney dataset: Normal–cyst–tumour–stone," Kaggle, 2023. [Online]. Available:

<https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone>

- [12] A. K. Kandepalli and A. G. Mote, "A deep convolutional neural network for high-accuracy classification of kidney cysts, tumours and stones in CT images," *TIJER – Technix Int. J. Eng. Res.*, vol. 12, no. 03, Mar. 2025, ISSN 2349-9249.
- [13] B. J. Leibiger, "Random forests for high-dimensional biomarker data in disease prediction," *Bioinformatics*, vol. 38, no. 15, pp. 3765–3773, 2022.
- [14] H. Bingol, A. Yildirim, and B. Diri, "Explainable deep learning for automatic kidney CT classification," *PeerJ Comput. Sci.*, vol. 9, p. e1405, 2023.
- [15] G. Macin, A. Akgul, and R. Yurt, "KidneyNeXt: A lightweight grouped-convolution CNN for kidney disease screening," *J. Clin. Med.*, vol. 14, no. 3, 2025.
- [16] T. H. Nguyen, R. Patel, and J. Smith, "Federated learning for privacy-preserving genomic data analysis across clinical sites," *npj Digit. Med.*, vol. 7, p. 44, 2024.
- [17] J. Park, S. Kim, and H. Lee, "Automated laboratory report parsing using OCR and named entity recognition for electronic health record population," *J. Biomed. Inform.*, vol. 134, p. 104191, 2022.