



الجمهورية العربية السورية
الجامعة السورية الخاصة
كلية هندسة الذكاء الاصطناعي
المشروع الفصلي

Intelligent Medical Diagnosis Assistant

مساعد تشخيص طبي ذكي

إشراف :

الدكتور: آصف جعفر

إعداد الطلاب:

حسين لؤي يوسف

أيمن عزت الحوراني

2026/2025

مقدمة:

يهدف هذا المشروع إلى بناء مساعد ذكي لدعم القرار الطبي يعتمد على منهجية Retrieval-Augmented Generation (RAG) مع نماذج تنبؤ إحصائية، وذلك لتقديم إجابات طبية مبنية على الأدلة وتقليل مشكلة هلوسة نماذج اللغة عند التعامل مع معلومات سريرية حساسة. يستقبل النظام وصفاً نصياً للحالة السريرية (case text) إضافةً إلى ميزات منظمة للمريض (مثل العمر، ضغط الدم، أمراض القلب، مؤشر كتلة الجسم، مستوى الغلوكوز، حالة التدخين...).

يتضمن النظام مسارين رئيسيين:

مسار التنبؤ بالمخاطر: يتم استخدام نموذج Random Forest لتقدير احتمال الإصابة بالسكتة الدماغية بالاعتماد على بيانات سريرية غير متوازنة (عدد الحالات الإيجابية قليل مقارنةً بالإجمالية)، لذلك تم اعتماد $\text{class weight}=\text{balanced}$ وضبط عتبة قرار منخفضة ($\text{threshold}\approx 0.05$) لرفع الحساسية وتقليل حالات الإغفال الطبي (False Negatives). كما يتضمن المشروع نموذجاً لتقدير خطر مرض القلب اعتماداً على ميزات سريرية من بيانات القلب (Cleveland)، مع إرجاع احتمالات التصنيف وتسميات مستوى الخطر.

مسار RAG الطبي: يتم بناء قاعدة معرفة من ملخصات PubMed الطبية، ثم تقسيم النص إلى مقاطع (chunking) وتحويلها إلى Embeddings وفهرستها عبر FAISS (HNSW) لاسترجاع المقاطع الأكثر صلة بالحالة. بعدها يُؤلّد النظام ملخصاً عربياً منظماً يربط عوامل الخطر بالأدلة المسترجعة ويعرض مصادر الاسترجاع ومستوى الثقة.

يضيف النظام آليات حماية وFallback عند ضعف الاسترجاع أو اضطراب خرج نموذج اللغة، ويُقدّم كمساعد داعم للطبيب وليس بديلاً عنه.

Abstract

This project develops an intelligent medical decision-support assistant that combines classical predictive modeling with **Retrieval-Augmented Generation (RAG)** to provide **evidence-grounded** medical outputs and mitigate language-model hallucinations in clinical contexts. The system accepts a free-text clinical description (**case text**) alongside structured patient features (e.g., age, hypertension, heart disease, BMI, average glucose level, smoking status, etc.).

The pipeline has two core tracks:

1. **Risk Prediction Track:** a **Random Forest** model estimates **stroke risk probability** from an imbalanced clinical dataset. To improve sensitivity to positive cases, the model uses **class weighting** and a tuned **low decision threshold (≈ 0.05)**, prioritizing the reduction of false negatives. In addition, a **heart disease risk** module (based on the Cleveland heart dataset features) outputs prediction probabilities and an interpretable risk label.
2. **RAG Evidence Track:** a medical knowledge base is built from relevant PubMed abstracts, chunked into passages, embedded into vector representations, and indexed using **FAISS (HNSW)** for efficient similarity search. For each case, the retriever returns the most relevant evidence snippets, which are then used to generate a structured Arabic medical summary grounded in retrieved evidence, including sources and a confidence signal.

The system also includes guardrails and fallback mechanisms for low-confidence retrieval or unstable LLM outputs. Overall, the project aims to support clinicians with evidence-based assistance rather than replace professional medical judgment.

الفصل الأول

المقدمة والإطار

العام للمشروع

1.1 مقدمة عن المشروع:

يشهد القطاع الصحي توسعًا متسارعًا في توظيف تقنيات الذكاء الاصطناعي (Artificial Intelligence) لدعم القرار السريري، خصوصًا في الحالات الحساسة التي تتطلب تدخلًا سريعًا ودقة عالية. يقدم هذا المشروع مساعدًا ذكيًا للتشخيص/الدعم الطبي يعتمد على دمج نماذج التعلم الآلي (Machine Learning) للتنبؤ بالمخاطر مع منهجية الاسترجاع المعزز بالتوليد (Retrieval-Augmented Generation – RAG) لتقديم إجابات أكثر موثوقية مدعومة بمقاطع معرفية مسترجعة من مصادر طبية. يركز النظام على التعامل مع حالة المريض عبر مدخلين: وصف نصي للحالة (case_text) وسمات سريرية منظمة (Structured Features)، ثم إنتاج مخرجات تتضمن تقدير مستوى الخطر وتفسيرًا مبنيًا على أدلة، مع مراعاة ضوابط السلامة الطبية.

1.3 الهدف من المشروع:

الهدف الرئيس هو تصميم وتنفيذ نظام يدعم القرار الطبي عبر:

- تقدير مخاطر الإصابة بالسكتة الدماغية باستخدام نموذج تعلم آلي مناسب للبيانات السريرية.
- توليد شرح طبي منظم ومدعوم بالأدلة باستخدام RAG استرجاع + توليد.
- رفع موثوقية المخرجات عبر إظهار "سببية/تفسير" قائم على مصادر مسترجعة بدل إجابة إنشائية غير موثقة.
- تحسين الحساسية تجاه الحالات الإيجابية عبر معالجة عدم التوازن وضبط عتبة القرار (Decision Threshold) عند الحاجة.
- تقديم النظام كأداة مساعدة لا كبديل عن الطبيب، مع ضوابط أمان و Fallback عند انخفاض الثقة.

1.2 المشاكل التي يقوم المشروع بحلّها:

يعالج المشروع مجموعة من التحديات الواقعية في تطبيقات الذكاء الاصطناعي الطبية، وأهمها:

1. **هلوسة نماذج اللغة: (LLM Hallucination)** قد تُنتج نماذج اللغة إجابات تبدو مقنعة لكنها غير صحيحة طبيًا عند غياب مرجعية واضحة.

2. **غياب الإسناد العلمي:** كثير من الأنظمة التوليدية لا تربط الإجابة بمصدر/دليل، ما يقلل الثقة والاعتمادية.

3. **تعقيد تفسير عوامل الخطر:** البيانات السريرية تتضمن عوامل متداخلة (Age, Hypertension, BMI, Glucose...)، ويصعب على المستخدم غير المختص فهم أثر كل عامل.

4. **عدم توازن البيانات: (Class Imbalance)** في بيانات السكتة الدماغية عادة تكون الحالات الإيجابية أقل بكثير من السلبية، ما قد يؤدي لانحياز النماذج نحو التنبؤ السلبي.

5. الحاجة إلى استرجاع معرفة متخصصة: الاعتماد على النموذج وحده دون قاعدة معرفة طبية يعرض النتائج للخطأ، بينما يتيح RAG ربط الإجابة بمقاطع معرفية ذات صلة.

6. السلامة: (Safety) يجب أن يتجنب النظام تقديم توصيات علاجية حاسمة أو قرارات طبية نهائية دون إشراف مختص.

1.4 مقدمة عن تجربة السكتة الدماغية (Stroke Experiment)

تُعد السكتة الدماغية من الحالات الحرجة التي تتأثر بعوامل خطر متعددة (عمر، ضغط، سكري/غلوكوز، أمراض قلب، نمط حياة...). في هذا المشروع تم بناء تجربة تهدف إلى تقدير مستوى خطر السكتة الدماغية اعتماداً على سمات سريرية منظمة، ثم ربط النتائج بتفسير مبني على الأدلة عبر RAG. تسير التجربة بمسارين متكاملين:

• **مسار تنبؤ: (Predictive Track)** يقدم قيمة احتمالية/تصنيف مستوى الخطر باستخدام ML.

• **مسار أدلة: (Evidence Track via RAG)** يسترجع مقاطع طبية مرتبطة بعوامل الخطر الظاهرة في الحالة ثم يولد ملخصاً تفسيرياً.

الفصل الثاني

الدراسة المرجعية

2.1 مقدمة

تُظهر الدراسات الحديثة في مجال التشخيص الطبي الذكي أن الاعتماد على منهجية واحدة (مثل نموذج تنبؤي فقط أو نموذج لغوي فقط) لا يكفي لتحقيق مستوى عالٍ من الدقة والموثوقية والسلامة. لذلك اتجهت الأبحاث إلى تطوير أطر هجينة تجمع بين: **استرجاع المعرفة الطبية** من مصادر موثوقة، و**تنظيم المعرفة** ضمن بنى معرفية (Knowledge Graph)، و**التفاعل متعدد المراحل/الجولات** لتحسين الاستدلال تدريجيًا كما يحدث في الممارسة السريرية.

بناءً على ذلك، يستعرض هذا الفصل ثلاث منهجيات مرجعية تُعدّ من الاتجاهات البارزة في "Medical AI" وهي:

1. **MedRAG (2025): RAG** مع رسم بياني معرفي تشخيصي هرمي.

2. **KG4Diagnosis (2024):** إطار تشخيصي قائم على Knowledge Graph مع بنية Multi-Agent.

3. **MRD-RAG (2025): RAG** متعدد الجولات (Multi-Round) لاستدلال تشخيصي تفاعلي.

4. ثم يوضّح الفصل آلية الدمج بين هذه المنهجيات لإنتاج فائدة جديدة في أنظمة الذكاء الاصطناعي الطبية.

2.2 المنهجية 1 — MedRAG (2025): الاسترجاع المعزّز بالتوليد ضمن هيكل معرفي هرمي

2.2.1 الفكرة الأساسية

يرتكز MedRAG على مفهوم **Retrieval-Augmented Generation (RAG)** بهدف رفع موثوقية الإجابات الطبية عبر ربط التوليد بمصادر معرفية مسترجعة. ويتميّز بأنه يدمج RAG داخل **هيكل معرفي طبي هرمي** (Hierarchical Diagnostic Knowledge Graph) يربط الأعراض بالأمراض والعوامل السريرية، مما يساعد على توجيه الاسترجاع وصياغة التفسير.

2.2.2 المكونات/الآليات

يتضمن MedRAG عادةً المكونات التالية:

- **استرجاع نصي أولي: (BM25)** لاستحضار الوثائق أو المقاطع المرشحة اعتمادًا على الكلمات المفتاحية.

- **استرجاع متجهي: (FAISS)** لتحسين الاسترجاع الدلالي عبر Embeddings.

• نموذج لغوي كبير: (LLM) لتوليد إجابة تفسيرية بعد تغذيته بالمقاطع المسترجعة.

• Fusion/Ranking: لدمج نتائج الاسترجاع متعددة المصادر وترتيبها حسب الصلة/الثقة.

2.2.3 الأثر العلمي (القوة/القيود)

• نقاط القوة: تحسين قابلية التفسير عبر الربط الهرمي بين المفاهيم الطبية، وتقليل الهلوسة عبر الاستناد إلى أدلة مسترجعة.

• القيود: يعتمد الأداء على جودة قاعدة المعرفة وجودة الرسم المعرفي؛ كما أن دمج أكثر من آلية استرجاع وترتيب يزيد التعقيد الهندسي والحوسبي.

2.3 المنهجية 2 — KG4Diagnosis (2024): رسم بياني معرفي + بنية متعددة الوكلاء لمحاكاة سير العمل السريري

2.3.1 الفكرة الأساسية

تتجه KG4Diagnosis إلى جعل التشخيص أقرب للواقع عبر تمثيل المعرفة الطبية ضمن Knowledge Graph وربط ذلك بهندسة Multi-Agent أي أن النظام لا يعمل ككتلة واحدة، بل يوزع المهام على “وكلاء (Agents) لكل منهم دور: فرز أولي (Triage)، تحليل اختصاصي (Specialist)، وتنسيق/تجميع القرار. (Coordinator)

2.3.2 المكوّنات/الآليات

- **Knowledge Graph:** يمثل علاقات (Symptom → Disease → Risk Factor) ويزيد التنظيم والتفسير.
- **Hybrid Retrieval:** يجمع بين الاسترجاع النصي (مثل BM25) والمتجهي (Embeddings + FAISS).
- **LLM + Agents:** لتوليد أسئلة متابعة أو تفسير النتائج ضمن سير عمل منظم.
- **تنسيق القرار:** دمج مخرجات الوكلاء بشكل قابل للتتبع.

2.3.3 الأثر العلمي (القوة/القيود)

- **نقاط القوة:** تنظيم تشخيصي واضح، قابلية توسع عالية، وإمكانية نمذجة “خط سير سريري” متعدد المراحل.

- **القيود:** زيادة التعقيد عند البناء (تعريف الوكلاء/الأدوار/التنسيق)، والحساسية لمشكلات ربط البيانات الطبية داخل الرسم المعرفي.

2.4 المنهجية 3 RAG — MRD-RAG (2025): متعدد الجولات لاستدلال تشخيصي تفاعلي

2.4.1 الفكرة الأساسية

يفترض MRD-RAG أن التشخيص الحقيقي لا يتم غالبًا من سؤال واحد، بل عبر سلسلة من الاستفسارات وجمع معلومات إضافية. لذلك يعتمد على **Multi-Round Retrieval-Augmented Generation** حيث يتم:

- توليد/اقتراح أسئلة متابعة،

- إعادة صياغة الاستعلام (Query Reformulation) ،

- تكرار الاسترجاع (Iterative Retrieval) ،

حتى الوصول إلى قرار أكثر استقرارًا وأقل غموضًا.

2.4.2 المكونات/الآليات

- حلقة تكرارية: (Loop) استعلام → استرجاع → تفسير → تحديث الاستعلام.
- تحسين الاسترجاع بناءً على التغذية الراجعة: لتحسين اختيار الأدلة في كل جولة.
- Hybrid Retriever: دعم الاسترجاع النصي والمتجهي.
- LLM للحوار الطبي: لتوليد أسئلة دقيقة وصياغة تفسير تدريجي.

2.4.3 الأثر العلمي (القوة/القيود)

- نقاط القوة: تقليل عدم اليقين، تحسين الاسترجاع تدريجيًا، والوصول لنتائج أقرب لمنطق الطبيب في المقابلة السريرية.

- **القيود:** زيادة زمن الاستجابة وتعقيد التنفيذ، والحاجة إلى تصميم سياسات توقف/ثقة (Stopping/Confidence) لمنع الدوران غير المفيد.

2.5 ملخص الفائدة وآلية الدمج بين المنهجيات الثلاث لإنتاج فائدة جديدة

2.5.1 الفائدة من كل منهجية

- **MedRAG:** يوفر “ظهر معرفي” قوي (Knowledge Backbone) يربط التوليد بأدلة ويزيد التفسير عبر بنية معرفية.

- **KG4Diagnosis:** يضيف “طبقة تنظيم” (Coordination Layer) “عبر-Multi-Agent Workflow تحاكي مراحل التشخيص.

- **MRD-RAG:** يضيف “طبقة تفاعلية” (Interactive Layer) “عبر تعدد الجولات لتحسين الاسترجاع والاستدلال تدريجيًا.

2.5.2 عملية الدمج (Integration) المقترحة

يمكن دمج المنهجيات الثلاث ضمن نموذج واحد بثلاث طبقات:

1. طبقة المعرفة والاسترجاع: (MedRAG) قاعدة معرفة + استرجاع (BM25/FAISS) توليد مسنود.

2. طبقة التنسيق متعدد المهام: (KG4Diagnosis) وكلاء تشخيص/فرز/تفسير وتوحيد القرار.

3. طبقة التحسين التفاعلي: (MRD-RAG) تعدد الجولات لإعادة صياغة الاستعلام وطلب معلومات إضافية.

2.5.3 الفائدة الجديدة في الذكاء الاصطناعي الطبي:

الفائدة الجديدة الناتجة عن الدمج هي بناء نظام لا يكتفي بـ "إجابة"، بل يقدم:

• قرارًا أكثر موثوقية لأن التوليد مُسند لأدلة مسترجعة،

• قابلية تفسير أعلى لأن مسار القرار يمكن تتبعه

• (Knowledge Graph + Agents)

- تكيفاً أكبر مع الحالات الناقصة عبر الحوار متعدد الجولات، وهو ما يرفع قابلية الاستخدام في التشخيص السريري كأداة دعم قرار (CDSS) بشكل أقرب للمنطق الطبي الواقعي.

الفصل الثالث

تنفيذ النظام

(Implementation)

6.1 لمحة عامة عن النموذج المنفذ

تم تنفيذ النظام على شكل معمارية هجينة (Hybrid Architecture) تجمع بين:

1. نموذج تنبؤ إحصائي (ML) لتقدير خطر السكتة الدماغية من بيانات سريرية منظمة (Tabular).

2. مسار RAG (Retrieval-Augmented Generation) لتوليد تفسير طبي عربي مُسند إلى مقاطع معرفية مسترجعة من قاعدة معرفة نصية طبية، بهدف تقليل الهلوسة وزيادة الموثوقية.

مخرجات النظام النهائية تتضمن:

• احتمال الخطر + تصنيف مستوى الخطر (Risk Label) (Probability)

• شرح/ملخص طبي عربي يربط عوامل الخطر بالأدلة المسترجعة

• آلية سلامة (Guardrails/Fallback) عند انخفاض الثقة أو ضعف الأدلة

6.2 شرح نموذج السكتة الدماغية المنفذ (Stroke Risk Model)

6.2.1 مدخلات النموذج

يعتمد نموذج السكتة على سمات سريرية منظّمة، من أهمها:

- العمر age
- ارتفاع ضغط الدم hypertension
- أمراض القلب heart_disease
- متوسط مستوى الغلوكوز avg_glucose_level
- مؤشر كتلة الجسم bmi
- حالة التدخين smoking_status
- سمات إضافية (مثل الجنس، الحالة الاجتماعية، نوع العمل، مكان الإقامة...) بعد ترميزها

6.2.2 معالجة البيانات قبل التدريب (Preprocessing)

تم تنفيذ خطوات معالجة معيارية لضمان جودة التدريب:

- تعويض القيم المفقودة خصوصاً bmi بأسلوب إحصائي مناسب.
- حذف العمود غير التنبؤي مثل id.
- ترميز المتغيرات الفئوية (One-Hot Encoding) لجعلها مناسبة لنماذج التصنيف.
- تقسيم البيانات تدريب/اختبار مع Stratify للحفاظ على توزيع الفئات.
- اعتمد المشروع على مجموعة بيانات سكتة دماغية منظمّة (Tabular Dataset) تضم 5110 سجلاً بمتغير هدف ثنائي $\text{stroke} \in \{0,1\}$ ، مع وجود عدم توازن طبقي واضح لصالح الفئة السلبية. تم تنفيذ سلسلة معالجة مسبقة قياسية شملت استبعاد الأعمدة غير ذات الدلالة التنبؤية (مثل id، معالجة القيم المفقودة ولا سيما في bmi عبر تعويض إحصائي مناسب، ثم ترميز المتغيرات الفئوية باستخدام One-Hot Encoding، وأخيراً تقسيم البيانات إلى تدريب/اختبار مع اعتماد stratified split للحفاظ على التوزيع الطبقي للفئات. ولأسباب تتعلق بالسلامة السريرية وتقليل احتمال إغفال الحالات الإيجابية، تم تبني عتبة قرار منخفضة ($\text{threshold} = 0.05$) بما يرفع الحساسية (Sensitivity/Recall) ويقلل الأخطاء من النوع الثاني (False Negatives) حتى مع احتمال ارتفاع نسبي في الأخطاء من النوع الأول (False Positives).

- بالتوازي، تم بناء مسار **RAG** عبر إنشاء قاعدة معرفة نصية طبية مرتبطة بالسكتة الدماغية وعوامل خطرها؛ حيث جرى تنظيف النصوص ثم تقطيعها إلى مقاطع (**Chunks**) وتمثيلها باستخدام **Embeddings** وفهرستها ضمن **FAISS** لتحقيق استرجاع دلالي فعال لأفضل المقاطع (**Top-k**) بهدف إسناد التفسير الطبي بالأدلة وتقليل هلوسة النماذج التوليدية.
- تم تقييم نموذج التنبؤ بالاعتماد على مصفوفة الالتباس (**Confusion Matrix**) واستخراج المقاييس القياسية **Accuracy** و **Precision** و **Recall** و **Specificity** و **F1-score**، بالإضافة إلى **ROC-AUC** و **PR-AUC** نظرًا لطبيعة عدم التوازن.
- أما مسار **RAG** فتم تقييمه بمؤشرات استرجاع مثل **Recall@k** و **Precision@k**، وبمعايير نوعية لجودة التوليد تشمل **Faithfulness/Groundedness** وسلامة المخرجات.

6.2.3 النموذج المستخدم وآلية التعلم

تم اعتماد نموذج **Random Forest Classifier** لكونه مناسبًا للبيانات الطبية الجدولية، وقادرًا على التعامل مع العلاقات غير الخطية بين العوامل السريرية.

ولأن بيانات السكتة غير متوازنة (الحالات الإيجابية أقل بكثير من السلبية)، تم تطبيق:

- `class_weight = "balanced"` لمعالجة عدم توازن الفئات (**Imbalance**).
- استخدام احتمال التنبؤ بدل الاعتماد على تصنيف مباشر فقط.

6.2.4 عتبة القرار (Decision Threshold)

تم ضبط عتبة القرار على:

• **threshold = 0.05**

والهدف من ذلك هو رفع الحساسية (Recall/Sensitivity) وتقليل False Negatives (تفويت حالات سكتة محتملة)، وهو خيار عملي في السياقات الطبية حيث يكون إغفال الحالة أخطر من الاشتباه الزائد). زيادة (False Positives)

6.2.5 مخرجات نموذج السكتة

ينتج النموذج:

- risk_probability: قيمة احتمال الخطر
- risk_label: تصنيف مستوى الخطر (مثل منخفض/متوسط/مرتفع بحسب منطق المشروع)
- (بيانات مساعدة تُستخدم لاحقاً في التفسير ضمن) RAG مثل أبرز عوامل الخطر الظاهرة من السمات

6.3 نموذج/وحدة خطر مرض القلب (Heart Risk Module)

يتضمن النظام وحدة مستقلة لتقدير خطر مرض القلب اعتماداً على سمات سريرية من بيانات القلب. (Cleveland)

تعمل الوحدة بطريقة مشابهة لمسار السكتة من حيث:

- إدخال سمات منظمة

- إخراج probabilities و prediction
- تحويل الناتج إلى تصنيف خطر يُعرض للمستخدم

وُستخدم مخرجاتها ضمن النص التفسيري عند تفعيل “مسار القلب” في النظام.

6.4 شرح مسار RAG المنفذ (الاسترجاع المعزز بالتوليد)

6.4.1 الهدف من RAG داخل المشروع

الغرض من RAG هو:

- تزويد النظام بأدلة طبية مسترجعة بدل توليد إجابة “من الذاكرة”
- تقليل الهلوسة
- رفع موثوقية الشرح الطبي عبر الاستناد إلى مقاطع من قاعدة معرفة

6.4.2 بناء قاعدة المعرفة النصية (Knowledge Base)

تم إعداد قاعدة المعرفة من نصوص/ملخصات PubMed طبية مرتبطة بالسكتة الدماغية وعوامل خطرها، ثم:

- تم وضع البيانات في abstracts
- تنظيف النصوص
- إزالة الضجيج قدر الإمكان
- تقسيمها إلى مقاطع قصيرة Chunks لتناسب طول السياق

6.4.3 تمثيل المقاطع Embeddings

بعد تجهيز الـ Chunks يتم تحويل كل مقطع إلى متجه (Embedding) باستخدام نموذج تمثيل نصي عميق (Transformer Encoder). هذا التمثيل يسمح بمقارنة "معنى" النص وليس الكلمات فقط.

6.4.4 الفهرسة والاسترجاع (FAISS Index)

تم بناء فهرس متجهي باستخدام FAISS بنمط HNSW لضمان:

- سرعة البحث
- قابلية التوسع
- استرجاع أفضل المقاطع ضمن Top-k

عند إدخال حالة جديدة:

1. يتم تحويل وصف الحالة/السؤال إلى Embedding
2. يُسترجع Top-k من المقاطع الأكثر تشابهًا دلاليًا
3. تُجهز المقاطع المسترجعة لإدخالها في مرحلة التوليد

6.4.5 التوليد المقيّد بالأدلة (LLM Generation)

يتم تمرير:

- ملخص الحالة (case_text)
- نتائج نموذج الخطر (الاحتمال/التصنيف)
- المقاطع المسترجعة (Evidence)

ضمن قالب Prompt منظم يفرض:

- تفسير عوامل الخطر
- ربط كل نقطة بما ورد في الأدلة
- تجنب الجزم الطبي/العلاجي

6.5 دمج المسارات داخل النظام (Integration Workflow)

آلية التنفيذ التشغيلية (Runtime Flow) تسير كالتالي:

1. استقبال مدخلات المستخدم (نص الحالة + السمات المنظمة)
2. تشغيل نموذج التنبؤ (السكتة/القلب) لإنتاج احتمال + تصنيف
3. تشكيل استعلام (Query) اعتمادًا على عوامل الخطر الظاهرة في الحالة
4. استرجاع أدلة Top-k من FAISS
5. توليد الشرح العربي باستخدام LLM اعتمادًا على الأدلة
6. تفعيل Guardrails/Fallback إن كانت الأدلة ضعيفة أو الثقة منخفضة
7. إخراج النتيجة النهائية بشكل منظم

6.6 آليات السلامة والتحكم (Guardrails & Fallback)

تم تضمين آليات لضمان السلامة، أهمها:

- عند ضعف الاسترجاع أو عدم كفاية الأدلة: يتم إخراج تفسير أكثر تحفظاً مع توصية بمراجعة مختص
- منع توصيات علاجية قطعية أو وصف أدوية
- ضبط أسلوب التوليد (مثل تقليل العشوائية) لتفادي نص غير منضبط

6.7 الأدوات المستخدمة (Tools & Technologies)

6.7.1 لغة البرمجة وبيئة التنفيذ:

- **Python 3.x**
- بيئة افتراضية لإدارة الاعتماديات + (venv/conda) ملف requirements

6.7.2 مكتبات معالجة البيانات وبناء النماذج

- **pandas:** قراءة وتنظيف وتحويل البيانات
- **numpy:** عمليات عددية ومعالجات مصفوفية
- **scikit-learn:** تدريب RandomForest ، تقسيم البيانات، المقاييس (Confusion Matrix, F1, ROC-AUC, ...)

6.7.3 مكتبات RAG والاسترجاع المتجهي

- **sentence-transformers:** توليد Embeddings للنصوص
- **FAISS:** بناء فهرس البحث المتجهي (HNSW) وتنفيذ Top-k retrieval

6.7.4 أدوات مساعدة

- مكتبات Python القياسية مثل `json`, `os`, `re`: لإدارة الملفات والنصوص والـ `metadata`
- (في حال تشغيل واجهة/خدمة) إطار API مثل **FastAPI** لتقديم النظام للمستخدم بشكل منظم

الفصل الرابع

الخلاصة والنتائج

والعمل المستقبلي

9.1 خلاصة بسيطة عن المشروع:

قدّم هذا المشروع نظامًا مساعدًا لدعم القرار الطبي يعتمد معمارية هجينة تجمع بين نموذج تنبؤ إحصائي لتقدير خطر السكتة الدماغية من بيانات سريرية منظمة، وبين مسار الاسترجاع المعزّز بالتوليد (RAG) لإنتاج تفسير عربي مُسند بالأدلة من قاعدة معرفة نصّية طبية. يركّز النظام على تقليل هلوسة النماذج التوليدية ورفع موثوقية الشرح، مع اعتماد قرار تشغيلي مناسب طبيًا يتمثل في ضبط عتبة القرار ($\text{threshold} = 0.05$) لتقليل تفويت الحالات الإيجابية.

9.2 النتائج (Results)

أظهرت النتائج العملية أن تبني معالجة عدم توازن البيانات وتغليب الحساسية عبر عتبة قرار منخفضة (0.05) يؤدي إلى:

- . تحسن واضح في الحساسية (Recall/Sensitivity) وتقليل حالات الإغفال الطبي للحالات الإيجابية. (False Negatives)
- . زيادة متوقعة في الإنذارات الكاذبة (False Positives) كنتيجة مباشرة
- . لخفض العتبة، وهو أثر مقبول في السيناريوهات التي تكون فيها كلفة الإغفال أعلى من كلفة الاشتباه.

أما في مسار RAG ، فقد ساهم الاسترجاع المتجهي وإدخال الأدلة ضمن التوليد في:

- . رفع اتساق الشرح وربطه بعوامل الخطر.
- . تقليل الميل إلى إنتاج معلومات غير مسنودة، بشرط توفر أدلة كافية ضمن قاعدة المعرفة وجودة استرجاع جيدة.

9.3 تفسير النتائج من منظور هندسي

يمكن تفسير سلوك النظام ضمن إطار تصميم أنظمة التصنيف غير المتوازنة (Imbalanced Classification) على النحو التالي:

1. **نقطة التشغيل (Operating Point) والتحكم بالمفاضلة**
خفض العتبة إلى 0.05 ينقل نقطة التشغيل نحو منطقة ترفع التقاط الحالات الإيجابية، أي زيادة Recall وتقليل FN. هندسيًا، هذا يعادل اختيار نقطة تشغيل تفضل تقليل الخطأ من النوع الثاني على حساب ارتفاع الخطأ من النوع الأول، وهو قرار منطقي عندما تكون كلفة FN أعلى طبياً وتشغيلياً.

2. **اعتمادية التفسير في RAG كسلسلة معالجة (Pipeline Reliability)**
موثوقية الشرح ليست خاصية للنموذج التوليدي وحده، بل هي ناتج لسلسلة معالجة متكاملة:
تمثيل دلالي → (Embeddings) فهرسة → (FAISS) استرجاع → (Top-k)
توليد مقيد بالأدلة.
أي ضعف في أي حلقة (مثل أدلة غير كافية أو استرجاع غير دقيق) سينعكس مباشرة على جودة التفسير، ما يفسر حساسية الأداء لتغطية قاعدة المعرفة وإعدادات الاسترجاع.

9.4 التحسينات المقترحة والمعالجة (Proposed Improvements)

9.4.1 تحسين توزع البيانات (Improving Data Distribution)

لتقليل أثر عدم التوازن وتحسين الأداء العام دون فقد الحساسية:

- تطبيق **Oversampling** للفئة النادرة (مثل SMOTE) أو **Undersampling** للفئة الأكبر، ثم مقارنة النتائج بمقاييس ملائمة مثل PR-AUC و F1.

- اعتماد **Stratified Cross-Validation** بدل تقسيم واحد (Train/Test) للحصول على تقدير أكثر ثباتًا للأداء وتقليل تحيز التقسيم.

- استخدام **Cost-Sensitive Learning** بشكل أوضح (أوزان/تكاليف أعلى للأخطاء السلبية) بما يتوافق مع الهدف الطبي.

- اختيار العتبة بطريقة منهجية عبر تحليل منحنيات Precision–Recall أو اعتماد شرط تشغيلي) مثل Recall أعلى من قيمة دنيا محددة (بدل اختيار ثابت غير مبرر).

9.4.2 استقرار التدريب والتعميم (Training Stability & Generalization)

لرفع الاستقرار وتقليل التذبذب بين التجارب:

- تثبيت **random_state** في جميع مراحل التقسيم والتدريب لضمان قابلية إعادة الإنتاج. (Reproducibility)

- ضبط معاملات النموذج عبر **Grid Search / Random Search** للحد من فرط التعلم (Overfitting) وتحسين التعميم، خصوصًا معاملات مثل :
`n_estimators, max_depth, min_samples_leaf.`

- تقييم النموذج باستخدام مقاييس مناسبة لعدم التوازن (PR-AUC ، Recall ، F1 بدل الاعتماد على Accuracy فقط.
- تحسين جودة البيانات عبر معالجة القيم الشاذة (Outliers) ووضع سياسة تعويض (Imputation) أكثر اتساقًا للقيم المفقودة.

9.4.3 (تحسين مسار RAG اختياري ضمن العمل المستقبلي)

- توسيع قاعدة المعرفة النصية بمواد أكثر حداثة وتغطية أوسع لعوامل الخطر.
- تحسين استراتيجية **Chunking** (حجم المقطع والتداخل) للوصول إلى مقاطع أكثر دلالة وأقل تشتيتًا.
- إضافة مرحلة **Reranking** بعد الاسترجاع لتحسين ترتيب الأدلة وتقليل إدخال مقاطع ضعيفة الصلة.
- تعريف مؤشر ثقة واضح لمسار الاسترجاع وتفعيل Fallback عند انخفاض التشابه/ضعف الأدلة.

الخاتمة

9.5 خاتمة:

يقدم هذا المشروع إطارًا عمليًا ومنهجيًا لبناء نظام دعم قرار طبي يجمع بين التنبؤ الإحصائي والتفسير المعرفي المُسند بالأدلة، وهو دمج يلبي حاجة أساسية في التطبيقات الطبية الحديثة: الحصول على قرار/تقدير مخاطر من جهة، والحصول على تفسير قابل للتتبع من جهة أخرى. وقد أثبتت التجربة أن الاعتماد على نموذج تنبؤ فقط even—لو كان جيد الأداء—يبقى محدودًا من ناحية الشرح والتبرير العلمي، بينما الاعتماد على نموذج لغوي توليدي وحده قد يعاني من ضعف الموثوقية بسبب احتمالية إنتاج معلومات غير دقيقة أو غير مسنودة. لذلك جاء التصميم الهجين ليوافق بين الدقة التشغيلية للنماذج الإحصائية وبين القدرة التفسيرية للمنهجيات التوليدية عند ربطها باسترجاع معرفي موثوق.

من منظور هندسي، يمكن النظر إلى النظام باعتباره سلسلة معالجة متعددة المراحل (Pipeline) لكل مرحلة وظيفة محددة ومخرجات واضحة. في مسار التنبؤ، تم التعامل مع الطبيعة غير المتوازنة لبيانات السكتة الدماغية بوصفها مشكلة تصنيف غير متوازن (Imbalanced Classification) تتطلب قرارات تشغيلية مختلفة عن التطبيقات العامة. وعلى هذا الأساس تم اعتماد عتبة قرار منخفضة ($\text{threshold} = 0.05$) بهدف رفع الحساسية (Sensitivity/Recall) وتقليل الأخطاء من النوع الثاني (False Negatives)، بما يتوافق مع المتطلبات السريرية التي تعتبر تفويت الحالة الأخطر عادةً من إطلاق إنذار إضافي. كما أن اختيار العتبة يعكس مفهوم "نقطة التشغيل" (Operating Point) "في هندسة الأنظمة، حيث يتم ضبط النظام وفق كلفة الأخطاء وسياق الاستخدام وليس وفق مقياس واحد مثل Accuracy.

أما مسار RAG، فقد قدم طبقة معرفية داعمة ساهمت في تحويل المخرجات من رقم احتمالي إلى تفسير طبي منظم، وذلك عبر ربط التوليد بسياق نصي مسترجع من قاعدة معرفة. هذا المبدأ يرفع الموثوقية لأنه يجعل النص الناتج أقرب إلى "تلخيص أدلة" بدل أن يكون "إنشاء حر". ومع ذلك، أوضحت التجربة أن جودة هذا المسار تعتمد بشكل مباشر على جودة حلقاته التقنية: جودة بناء قاعدة المعرفة، كفاءة التقطيع (Chunking)، جودة تمثيل النصوص عبر Embeddings، ودقة الاسترجاع عبر FAISS، ثم جودة تشكيل المدخلات (Prompting) للنموذج التوليدي. وعليه فإن موثوقية التفسير هي ناتج تكاملي لهندسة النظام بالكامل، وليست خاصية منفصلة لنموذج واحد.

تؤكد نتائج المشروع أن الأنظمة الطبية الذكية الأكثر نضجًا هي تلك التي لا تكتفي بالإجابة، بل تركز على: سلامة القرار، قابلية تفسيره، وشفافية مصادره. لذلك تم تضمين آليات سلامة (Guardrails) واستجابات تحفظية (Fallback) عند ضعف الأدلة أو انخفاض الثقة، بما يدعم الاستخدام المسؤول للنظام في سياقات صحية حساسة. ومع أن النظام لا يُقدّم كبديل عن الطبيب، إلا أنه يبرهن قدرة ملموسة على دعم عملية التقييم الأولي وربط عوامل الخطر بالأدلة المتاحة، وهو ما يجعله مناسبًا كنواة لنظام دعم قرار سريري (CDSS) قابل للتطوير.

وفي ضوء ما سبق، تبرز أهمية العمل المستقبلي في محورين متكاملين: محور **تحسين البيانات والتعميم** عبر تقنيات معالجة عدم التوازن (مثل إعادة التوزيع والتقييم بالتقسيمات الطبقيّة المتعددة)، ومحور **تحسين جودة الاسترجاع والتفسير** عبر توسيع قاعدة المعرفة، وضبط إعدادات الاسترجاع، وإضافة مراحل إعادة ترتيب الأدلة (Reranking)، وربما اعتماد تقييم بشري متخصص للمخرجات. إن تطوير هذه المحاور من شأنه رفع دقة النظام وثباته، وتوسيع نطاقه ليغطي حالات سريرية أكثر مع الحفاظ على متطلبات السلامة والموثوقية.

وبذلك يحقق المشروع هدفه الأساسي في تقديم نموذج تطبيقي متكامل يبرهن أن دمج **ML** **للتنبؤ مع RAG للتفسير المسنود** يمكن أن ينتج نظامًا طبيًا أكثر اتزانًا بين الأداء والموثوقية، وأكثر قربًا من احتياجات الاستخدام السريري الواقعي، مع إمكانية تطويره مستقبلاً ضمن إطار علمي وهندسي واضح.