

Use case and datasets

Use Case	Goal	Dataset Type	Example Features	Potential ML/AI Techniques	Data Quality Requirements	Possible Sources
1. Intelligent Claims Processing & Automation	Reduce claim processing time and improve accuracy by automating verification and classification	Historical claims database; medical reports; employer records	Claim type, injury severity, treatment plan, employer risk rating, processing time	Supervised classification (Random Forest, XGBoost), OCR + NLP for text extraction	Complete claim forms, standardized injury codes, digitized historical files, minimal missing fields	WCF claims system, affiliated hospitals, Ministry of Labour databases
2. Fraud Detection	Detect suspicious or fraudulent claims early to prevent losses	Labeled claims (fraud vs. non-fraud); claimant history; billing data; external datasets	Claim frequency, unusual billing amounts, claimant–employer links, accident timing patterns	Anomaly detection (Isolation Forest, Autoencoders), Graph-based fraud detection	Accurate fraud labels, time-stamped data, normalized billing codes, verified claimant IDs	WCF fraud investigation unit, NHIF, police reports, court records
3. Predictive Risk Assessment	Identify high-risk sectors, employers, or employees for targeted safety interventions	Employer risk profiles; workplace safety inspections; employee demographics	Accident history, industry type, compliance scores, age group, job category	Predictive modeling (Gradient Boosting, Decision Trees), Risk scoring	Regularly updated compliance scores, industry classification consistency, anonymized personal data	OSHA inspection reports, NBS labour statistics, WCF employer registry
4. Workplace Safety Monitoring	Monitor work environments to prevent accidents and improve safety compliance	IoT/sensor data; CCTV video feeds; PPE compliance logs; maintenance records	Temperature, noise, PPE usage, machine operating hours	Computer Vision (YOLO, Faster R-CNN), Time-series forecasting (LSTM, Prophet)	High-resolution video, synchronized sensor timestamps, real-time data streaming capability	Partnered factories, warehouses, IoT devices in workplaces, WCF–OSHA joint inspections
5. Customer Service Chatbots & Document Search	Provide instant and accurate responses to stakeholder queries	FAQ knowledge base; past customer service interactions; legal documents	Question–answer pairs, claim status logs, law text	NLP-based Q&A (BERT, GPT fine-tuning), Search engines	Structured FAQs, clean chat/email transcripts, tagged intents, up-to-date laws	WCF website, call center CRM, Government Gazette (Workers Compensation Act)
6. Medical Cost Forecasting	Estimate medical costs for claims to improve budgeting and fund allocation	Medical billing records; diagnosis & treatment data; recovery timelines; economic trends	Procedure codes, cost per treatment, inflation rates, return-to-work dates	Regression models (XGBoost Regressor), Time-series forecasting	Accurate cost coding, consistent diagnostic categories, inflation index alignment	WCF medical panel, NHIF, Ministry of Health, Bank of Tanzania economic data
7. Investment & Fund Management Optimization	Maximize investment returns while ensuring liquidity for payouts	Historical investment performance; market data; cash flow records	Asset type, ROI, volatility, premium inflow, payout history	Portfolio optimization (Markowitz, RL), Predictive forecasting	Consistent financial records, asset classification, real-time market feeds	WCF Finance Dept., Dar es Salaam Stock Exchange, Bank of Tanzania, Bloomberg/Refinitiv feeds