6COSC023W

Computer Science Final Project

Project Proposal and Requirements Specification (PPRS)

Handout

**A sound recognition system for monitoring safety at workplaces**

**(SoundMonitor AI)**

**Student**: Dat Quang Pham (20992053)

**Supervisor**: Bob Nguyen

**Degree:** BSc (Hons) Computer Science

School of Computer Science & Engineering

College of Design, Creative and Digital Industries

University of Westminster

13-11-2025

# Declaration

This document has been prepared based on my own work. Where other published and unpublished source materials have been used, these have been acknowledged in references.

Student Name: Dat Quang Pham

Date of Submission: 13-11-2025

# Definition of the Problem

* **Problem Description**:

Workplace safety remains a critical concern across UK industries, particularly in high-risk sectors such as construction, agriculture, and manufacturing. According to the Health and Safety Executive (HSE, 2023a; HSE, 2023b), there were 604,000 non-fatal injuries in 2023/2024 and 124 fatal injuries in 2024/2025, costing the UK economy over £20 billion annually through direct costs (medical treatment, compensation) and indirect costs (production delays, insurance premiums, legal proceedings).

Current workplace monitoring systems predominantly rely on manual supervision by safety officers and CCTV camera surveillance. However, these approaches suffer from critical limitations: manual supervision cannot continuously monitor all areas simultaneously, while CCTV systems have inherent blind spots in complex industrial environments and require constant human attention to identify developing hazards. Neither approach can detect acoustic warning signs—equipment malfunctions producing abnormal sounds (grinding bearings, overheating motors), structural failures generating stress sounds (creaking scaffolds, cracking materials), or worker distress calls in noisy environments—that often precede visible incidents by minutes to hours, resulting in delayed emergency response.

* **Significance**:

This project addresses these limitations through an automated acoustic monitoring system that detects workplace hazards via sound event classification. The proposed system employs deep learning-based audio analysis to identify five critical hazard categories: equipment malfunctions (grinding bearings, overheating motors), falling objects and collisions (impact sounds), structural failures (creaking scaffolds, material stress), worker distress signals (calls for help in high-noise environments), and emergency alarms (fire alarms, evacuation signals). Upon detection, the system triggers multi-channel alerts (visual, audio, network notifications) to enable rapid intervention within the critical first 60 seconds when outcomes remain modifiable

# Aims, Objectives, and Scope

* **Aim**:

Design and implement a real-time workplace safety monitoring system using sound event detection to identify dangerous workplace sounds including equipment malfunctions, falling objects, structural failures, worker distress signals, and emergency alarm and deliver immediate alerts to enable rapid intervention.

**Objectives**:

- Function: Design and implement a real-time workplace safety monitoring system using sound event detection to identify dangerous workplace sounds including equipment malfunctions, falling objects, structural failures, worker distress signals, and emergency alarms and deliver immediate alerts to enable rapid intervention in high-risk environments such as construction sites, manufacturing facilities, and warehousing operations.

- Performance Targets: Develop and train a deep learning model for accurate classification of workplace hazard sounds using transfer learning from pretrained audio networks.

- Demonstrate system scalability by optimizing the trained model for deployment on low-cost edge computing hardware following successful desktop implementation.

* **Scope**:
  + **Inclusions**:
    - Real-time audio classification in constrained workplace environments (office buildings, warehouses, manufacturing facilities)
    - System can detect and recognise hazardous acoustic events such as:

Emergency alarms and sirens

* + - * Human distress sounds (shouting, screaming, calls for help)
      * Object collision and impacts (falling objects, crashes)
      * Glass breaking and structural failures
    - Continuous audio monitoring with event logging and timestamping
    - Daily sound event recording and storage for historical analysis and incident investigation
    - Alert notification system (visual, audio, network-based notifications to monitoring department)
    - Deployment limited to indoor workplace environments with controlled acoustic conditions
  + **Exclusions**:
    - Environment limitations:
      * Outdoor environment deployment (open-air workplaces, construction sites, outdoor events)
      * Extremely high-noise environments (factories with >85dB sustained noise, stadiums, concerts, outdoor festivals)
      * Variable weather conditions (rain, wind, thunder affecting outdoor audio quality)
    - **Acoustic Event Limitations:**
      * Fire/flames in early stages (crackling sounds often <40dB, masked by ambient noise)
      * Gas leaks (inaudible or ultrasonic frequencies beyond detection range)
      * Slow structural degradation (gradual cracking, imperceptible acoustic changes)
      * Chemical reactions without audible manifestation

# Background Review

* **Key Findings (Literature / Systems)**:

The literature review conducted for this study identified several key features that are essential to robust sound event detection systems:

Contemporary audio classification research demonstrates a clear trend toward CNN-based architectures (Hershey, S et al., 2017. Kong, Q., Cao et al., 2020. Tsalera, E. et al., 2021). CNNs excel at extracting high-level features invariant to local spectral and temporal variations, while RNNs capture longer-term temporal dependencies within audio signals. Recent work has shown that hybrid CRNN architectures, combining the spatial feature extraction capabilities of CNNs with RNN temporal modelling, achieve significant performance improvements, absolute gains of 6.6% and 13.6% in frame-based F1 scores compared to CNN and RNN, respectively approaches (Cakir et al., 2017). This finding motivates the proposed system's adoption of CRNN architecture for workplace sound event detection.

Audio preprocessing employs log-mel spectrograms, a 2D time-frequency representation that transforms 1D waveforms into 2D matrices compatible with CNN processing. Comparative studies demonstrate that audio-specific CNN architectures (e.g., VGGish, YAMNet) pre-trained on large-scale audio datasets substantially outperform image-based CNNs (e.g., ImageNet pre-trained models) for audio classification tasks (Tsalera, Papadakis and Samarakou, 2021). Tsalera et al. report that sound CNNs achieve 96.7% accuracy on UrbanSound8K with 80% faster training time compared to image CNNs requiring extensive data augmentation (Tsalera, Papadakis and Samarakou, 2021). This evidence supports the selection of audio-specific pre-trained models (PANNs, YAMNet) as transfer learning baselines for this project.

Existing audio classification research relies on generic environmental datasets: YouTube-100M (30,871 labels) (Hershey et al., 2017), AudioSet (527 classes) (Kong et al., 2020), UrbanSound8K (urban sounds) (Mu et al., 2021), and specialized single-equipment datasets (air compressor) (Tsalera, Papadakis and Samarakou, 2021). However, these datasets lack representation of workplace safety-critical sounds equipment malfunctions producing abnormal acoustic signatures (grinding bearings, overheating motors), structural failures generating warning sounds (creaking scaffolds, cracking materials), and worker distress signals (calls for help) obscured by industrial ambient noise (>80dB). This domain gap is significant: HSE UK reports 604,000 non-fatal workplace injuries annually (HSE, 2023b), many preceded by detectable acoustic signatures, yet no public audio dataset specifically targets workplace operational hazards. This project addresses the gap through two approaches: (1) mapping existing dataset classes to workplace-relevant proxies (e.g., gunshot → collision, siren → alarm), and (2) collecting custom workplace-specific audio samples during deployment for domain-specific fine-tuning. The resulting system focuses on five hazard categories absent from existing research: equipment malfunction, falling objects/collision, glass breaking/structural failure, human distress, and emergency alarms.

Transfer learning has emerged as a critical technique for reducing computational requirements in audio classification. Tsalera et al. demonstrate that fine-tuning pre-trained audio models reduces training time by approximately 80% compared to training from scratch, while maintaining or exceeding baseline accuracy (Tsalera, Papadakis and Samarakou, 2021). This efficiency enables rapid experimentation and reduces GPU infrastructure requirements, particularly beneficial for resource-constrained research environments.

State-of-the-art pre-trained audio models achieve strong benchmark performance. Kong et al.'s PANNs system attains mean Average Precision (mAP) of 0.439 on AudioSet's 527-class taxonomy (Hershey et al., 2017), while audio-specific architectures (VGGish, YAMNet) demonstrate superior transfer capabilities compared to image-based models for audio classification tasks (Tsalera, Papadakis and Samarakou, 2021). These models provide robust feature extractors for downstream applications with limited labeled training data.

Workplace safety hazards, by their nature, constitute rare events in normal operational data distributions, yet the literature predominantly evaluates models on balanced or near-balanced datasets. The challenge of maintaining high recall for infrequent but critical events (e.g., structural failures, emergency alarms) while controlling false positive rates on vastly more common background sounds remain

* **Comparison with Similar Software Applications/Products**

Product 1: SONITROL Audio Intrusion Detection System (Sonitrol, 2025)

**Strengths:**

* + - Proven reliability with 187,000+ apprehensions: Human-verified alarms achieve lowest false alarm rate in industry (80% fewer than average), leading to 8-minute average police response
    - Comprehensive volumetric coverage: Wall-to-wall, floor-to-ceiling audio detection detects intrusions before entry (glass breaking, door prying), unlike motion sensors limited to specific points

**Weaknesses:**

* + - No AI/ML capabilities: Entirely human-dependent verification creates 8-minute delay; cannot autonomously classify sound events or operate without monitoring station connection
    - Limited to security intrusion only: Focused exclusively on break-ins/theft; does not detect workplace safety hazards (equipment failures, worker distress, structural issues)

Product 2: March Networks AI Series Smart Cameras with Audio Monitoring (March Networks, 2025)

**Strengths:**

* + - Integrated multi-modal surveillance: 4MP cameras combine video analytics, AI-powered people/vehicle detection, and audio monitoring in single device with no additional licensing fees
    - NDAA-compliant with edge processing: Built-in analytics run on-camera for real-time threat detection (gunshots, glass breaking, aggression) meeting government security standards

**Weaknesses:**

* + - Very high cost with cloud dependency: $500-1500+ per camera with best features requiring cloud subscription and constant internet; primarily video-centric design treats audio as secondary add-on feature
    - Not workplace safety-focused: Generic audio detection (gunshots/glass break) lacks training for industrial hazards; cannot detect equipment failures, worker distress signals, or operational anomalies

Product 3: IPVideo HALO Smart Sensor (HALO Smart Sensor, n.d.)

**Strengths:**

* + - AI-powered keyword and aggression analytics: Machine learning identifies threats through keyword alerts (bullying, help calls) and aggression detection via voice tone analysis in real-time

**Weaknesses:**

* + - Limited to education-focused keywords: Optimized for school environments (bullying, vaping); cannot classify diverse workplace hazards like equipment malfunctions, structural failures, or operational sounds
  + **Differentiation**:

1. Commercial Products

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **SONITROL** | **March Networks AI** | **IPVideo HALO** | **My Project** |
| **Primary Function** | Intrusion detection | Video surveillance + audio | Multi-hazard sensor | Workplace safety monitoring |
| **Technology** | Impact-activated audio sensors | AI camera + microphone | AI audio analytics + sensors | Deep learning sound classification |
| **Sound Detection Method** | Human verification | Threshold-based (dB levels) | ML keyword/aggression | CNN/CRNN trained on workplace sounds |
| **AI/ML Capabilities** | None | Limited (video AI, basic audio) | Yes | Yes |
| **Sound Event Types** | Glass break, door/window breach | Gunshots, glass break, aggression | Gunshots, aggression, keywords | Equipment failures, distress signals, alarms, collisions, glass break |
| **Real-time Classification** | No | No | Yes | Yes |
| **Deployment Platform** | Proprietary hardware + central station | Enterprise cameras + cloud | Proprietary sensor + cloud | Laptop/PC (Mandatory)  Raspberry Pi 4 (Optional) |
| **Hardware requirements** | Proprietary sensors | IP cameras + microphones | Proprietary sensor | Standard laptop or RPi 4 |
| **Installation Effort** | High (professional) | High (professional) | Medium (semi-pro) | Low(software install) |
| **Accuracy** | Human verification (high) | Not specified | Not specified | Target: >90% on workplace sounds |
| **Target Environment** | Security intrusion | Retail, banking, transit | Schools, public restrooms | Offices, warehouses, factories |

**Table 1. Commercial Products Comparison**

# A systematic evaluation of existing commercial workplace safety monitoring solutions has revealed several critical gaps and opportunities that substantiate the rationale for developing SoundMonitor AI. The most significant limitation observed across current market offerings is the absence of real-time classification capabilities in audio monitoring systems. The majority of commercial products operate with notable processing delays, which consequently result in delayed intervention when workplace incidents or accidents occur. This latency introduces a critical safety vulnerability, as timely response is paramount during emergency situations such as equipment malfunctions, worker injuries, or security threats. By implementing genuine real-time sound event detection, SoundMonitor AI seeks to enable instantaneous alerting mechanisms that can substantially reduce emergency response times and potentially mitigate serious injuries or fatalities.

# Furthermore, the commercial product analysis identified significant gaps in the scope of monitored audio events. While many existing solutions predominantly focus on industrial machinery acoustics or general anomaly detection, there remains insufficient coverage of critical safety scenarios including active shooter incidents and workplace violence. To address these shortcomings, the proposed system will incorporate gunshot detection and aggression-related acoustic patterns into the dataset collection methodology. This encompasses not only the acoustic signatures of firearms but also distress vocalizations and verbal altercations that may indicate escalating conflicts or emergency situations necessitating immediate security intervention. This expanded coverage positions SoundMonitor AI as a more comprehensive and holistic workplace safety solution.

# The hardware deployment strategies employed by contemporary commercial solutions present a substantial barrier to widespread adoption, particularly for small and medium-sized enterprises. Current market offerings are predominantly characterized by expensive proprietary edge devices, rendering these safety systems financially prohibitive for numerous businesses that would benefit significantly from such technology. SoundMonitor AI addresses this accessibility challenge through a phased deployment methodology. Initially, the system will be developed and optimized on standard laptop or personal computer platforms, which provides a more economically viable solution during the development and validation phases whilst simultaneously facilitating the debugging and refinement processes. Upon demonstrating reliable performance metrics on this foundational platform, the system will subsequently be migrated to Raspberry Pi 4 devices. These single-board computers are particularly appropriate for small to medium-scale business applications owing to their cost-effectiveness, compact architecture, and sufficient computational resources for real-time audio processing tasks.

# Finally, the proposed system distinguishes itself through the implementation of an automated analytics pipeline leveraging essential Python libraries including NumPy, Matplotlib, and scikit-learn. This technological framework enables the system to autonomously collect, process, and analyze acoustic data without necessitating continuous human supervision. The automated analytics capability not only alleviates the monitoring burden on personnel but also delivers valuable insights through pattern recognition algorithms and data visualization techniques. This functionality allows organizations to identify longitudinal trends in workplace safety incidents and formulate evidence-based decisions regarding risk mitigation strategies and resource allocation.

# Tools (Hardware/Software)

* **Hardware**:
  + **Laptop/PC**
  + **Raspberry Pi 4 Model B 8GB**

### **Primary Development Platform: Laptop/PC (Mandatory)**

The development workstation features a 13th Generation Intel Core i5-13450HX processor with 10 cores, providing substantial parallel processing capabilities essential for concurrent data preprocessing and model training operations. The 16GB DDR4 RAM enables efficient batch processing of audio datasets without memory bottlenecks, while the NVIDIA GeForce RTX 5060 GPU with 8GB VRAM accelerates CNN-LSTM training by approximately 10-15x compared to CPU-only implementations through its optimized Tensor Cores. The 1TB NVMe SSD delivers sequential read speeds exceeding 3000MB/s, minimizing data loading delays and maintaining high GPU utilization throughout training. The integrated microphone array facilitates immediate real-time testing without external recording equipment, while Windows 11's native CUDA 11.8 support ensures optimal deep learning framework performance.

**Deployment Platform: Raspberry Pi 4 Model B 8GB (Optional)**

The Raspberry Pi 4 Model B with 8GB RAM is recommended for edge deployment, featuring a Quad-core Cortex-A72 processor at 1.5GHz, dual-band wireless connectivity, and 40-pin GPIO for sensor integration. The 8GB memory variant provides sufficient capacity for loading optimized TensorFlow Lite models alongside audio processing libraries and buffering mechanisms required for real-time operation. With model optimization techniques such as quantization and pruning, the platform can achieve inference times of 100-300ms per audio segment, satisfying real-time workplace safety monitoring requirements. The device's compact form factor (85.6mm × 56.5mm), low power consumption (6-8W), and cost-effectiveness (approximately £70-80) make it economically viable for deploying multiple units across facility zones. The mature Raspberry Pi ecosystem offers extensive community support, well-documented TensorFlow Lite optimization guides, and proven audio processing deployment examples that reduce implementation risk.

**Software**:

**Development Environment:**

**VS Code, Jupyter Notebook:**

* + - Visual Studio Code provides comprehensive IDE capabilities with extensive extension support including Python IntelliSense, Pylance for type checking, and integrated Git functionality that streamlines development workflows. Its Remote Development extensions enable direct development on Raspberry Pi devices via SSH, facilitating deployment testing.
    - Jupyter Notebook complements VS Code by offering an interactive environment ideal for exploratory data analysis, algorithm prototyping, and inline visualization of spectrograms and training metrics. The cell-based execution model enables rapid experimentation without full script re-execution, while notebooks serve as executable documentation for reproducible experiments. Both environments together create an optimal workflow for iterative development, debugging, and clear communication of methodologies.

**Programming Language and Core Libraries:**

* + - **Python**:

Python was selected due to its dominance in machine learning with extensive audio processing library support, abundant community resources, and implementations of cutting-edge audio classification techniques.

* + - **TensorFlow, PyTorch:**

TensorFlow and PyTorch offer complementary advantages: PyTorch provides intuitive development with dynamic computational graphs for rapid prototyping, while TensorFlow excels in deployment through TensorFlow Lite's edge optimization and comprehensive quantization tools for Raspberry Pi.

* + - **Librosa, Pyaudio:**

Librosa serves as the standard library for audio feature extraction, providing efficient implementations of MFCCs, spectrograms, and Log-Mel transformations critical for CNN preprocessing. PyAudio enables real-time audio capture through its streaming interface and callback mechanism, ensuring low-latency continuous monitoring without blocking I/O operations essential for real-time detection.

**Data Analysis and Visualization Libraries:**

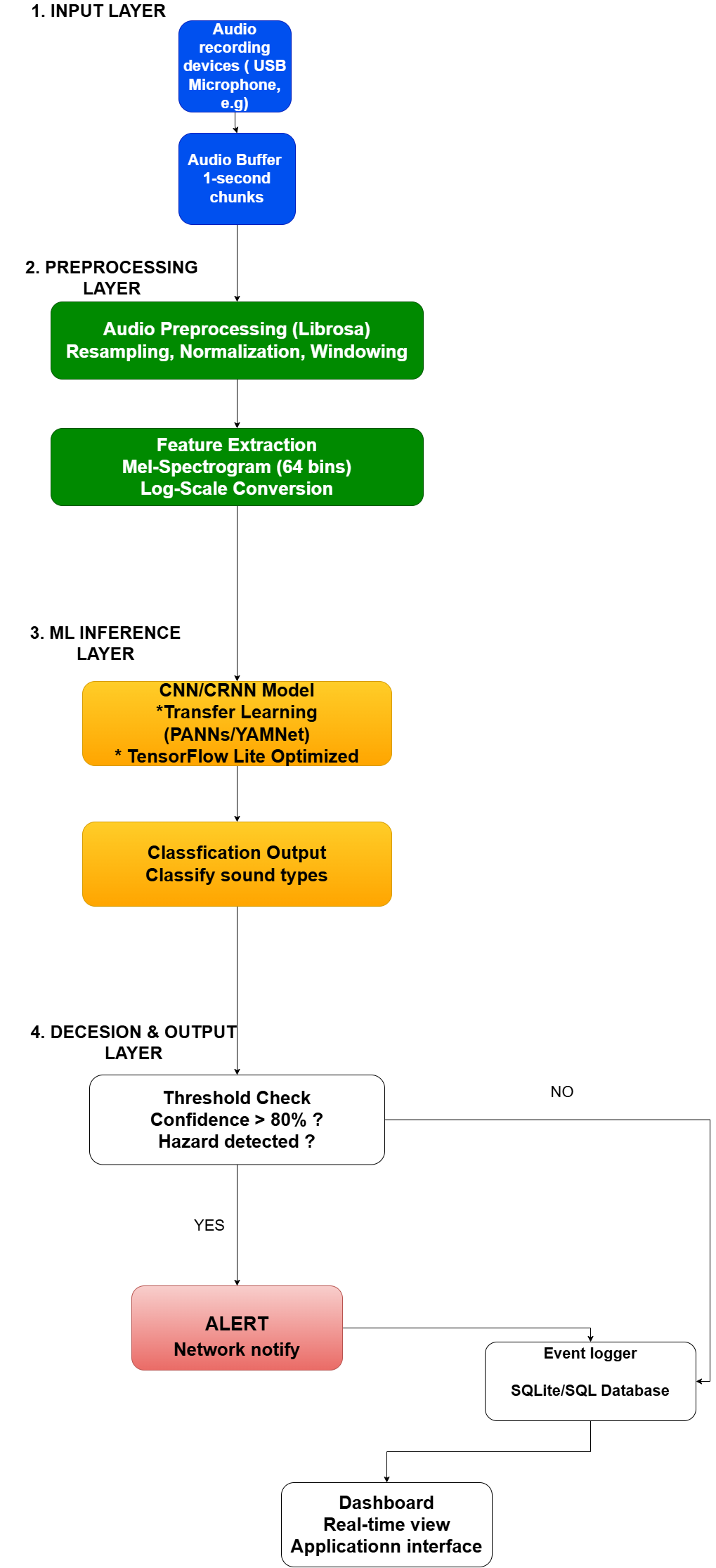
* + - **NumPy, SciPy, and Pandas** form the foundational stack for numerical computing and data manipulation, with NumPy providing vectorized array operations that execute orders of magnitude faster than pure Python for audio signal processing. SciPy extends these capabilities with advanced signal processing functions including filtering and spectral analysis necessary for preprocessing audio in noisy industrial environments. Pandas facilitates structured management of metadata, training logs, and performance metrics for systematic model behaviour analysis.
    - **Matplotlib and Seaborn** enable creation of publication-quality spectrograms, confusion matrices, and training curves essential for diagnosing issues and demonstrating system performance in project documentation.

**Version Control:**

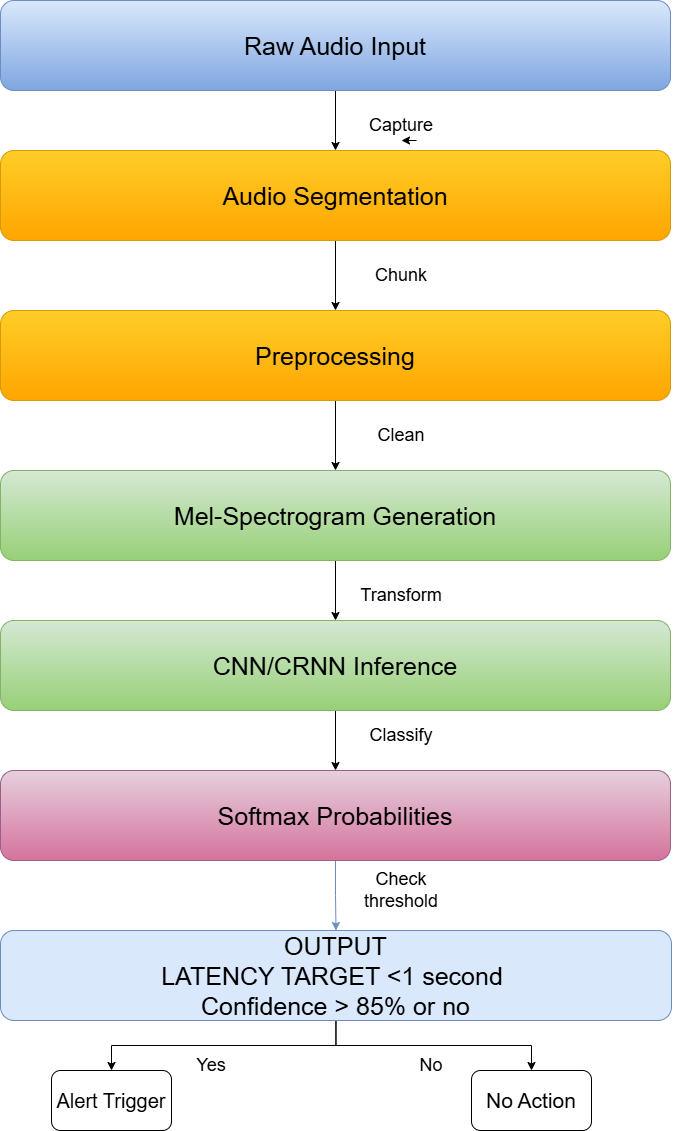
* + - **Github:**

GitHub provides comprehensive version control alongside issue tracking, project boards for agile workflow management, and CI/CD capabilities through GitHub Actions. GitHub serves dual purposes: maintaining professional development practices with Git workflows and functioning as a portfolio platform to demonstrate technical capabilities for future employment opportunities.

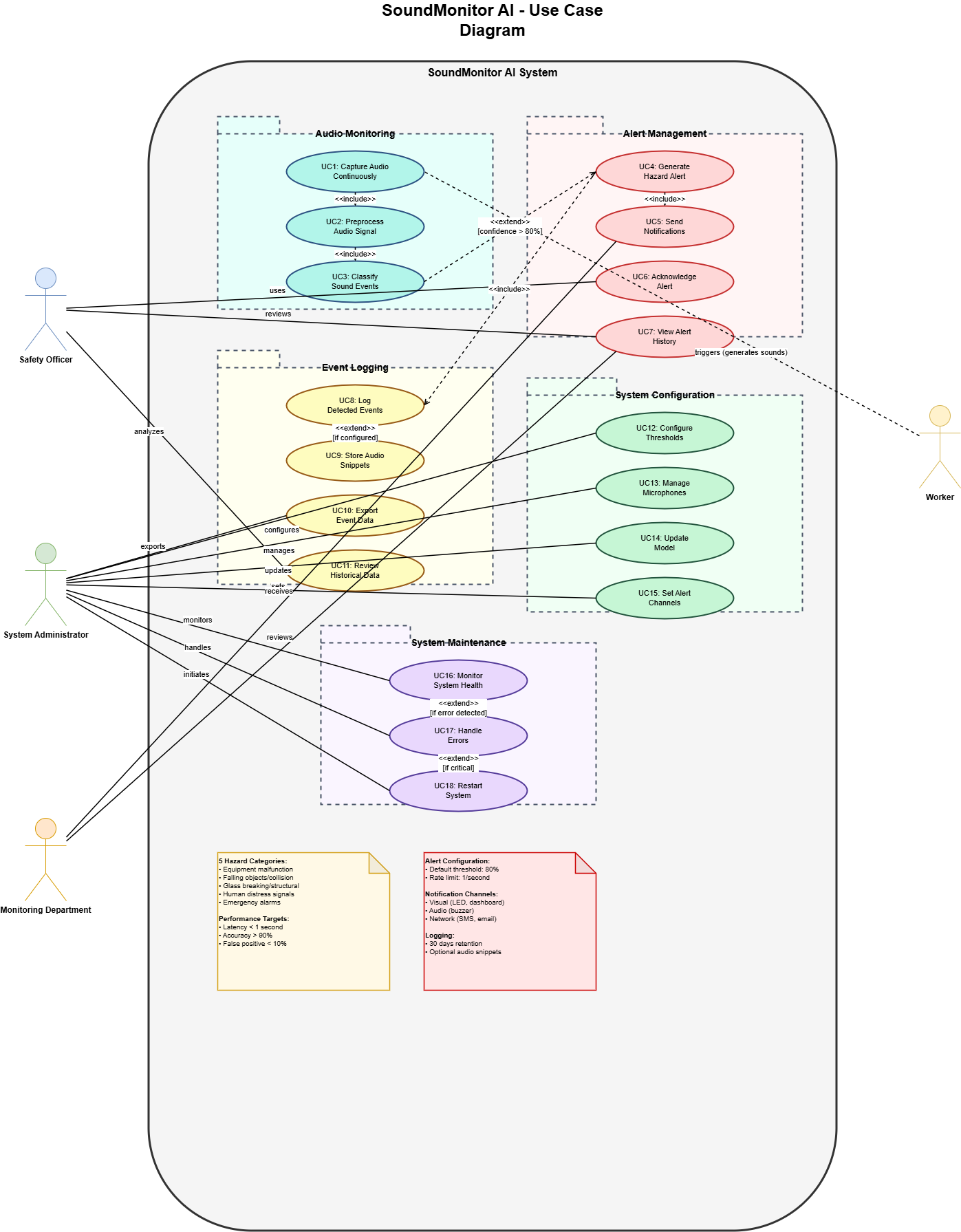
# Use Cases and Diagrams



**Diagram 1. System Architecture**

****

**Diagram 2. Data Flow**

****

**Diagram 3. Use Cases**

# Requirements Elicitation

* **Approach**:

**Method 1: Literature Analysis**

Literature reviews on many papers as (Hershey et al. 2017, Kong et al. 2020, Elangovan et al. 2021, Mu et al. 2021, Tsalera et al. 2021) established technical benchmarks and feasibility. This revealed state-of-the-art models achieve 87-95% accuracy on standard datasets, real-time inference requires <1 second latency, and transfer learning reduces training time by 80% while maintaining accuracy. These findings confirmed the technical viability of achieving >90% detection accuracy with <1 second latency on resource-constrained hardware.

**Method 2: Competitive Analysis**

Three commercial workplace safety products were evaluated (SONITROL, March Networks, IPVideo HALO) to identify market gaps and best practices. Key insights included cost barriers ($500-1500 per unit), privacy concerns with video surveillance, domain gaps (security-focused rather than workplace hazards), and cloud dependency issues. These limitations directly informed requirements for edge-based processing, audio-only monitoring, and <$150 cost target.

**Method 3: Regulatory Review**

UK workplace safety regulations and GDPR requirements were analyzed to ensure legal compliance. HSE UK statistics (565,000 annual injuries, £20 billion economic cost) established problem severity, while GDPR data protection regulations mandated privacy-preserving design requirements including on-device processing and no cloud audio uploads.

**Method 4: Public Dataset Collection and Analysis**

Public audio datasets were identified and evaluated to support model training and validation. Three primary datasets were selected: UrbanSound8K (8,732 labeled sound excerpts, 10 classes including glass breaking and sirens), ESC-50 (2,000 environmental sound clips covering 50 classes), and AudioSet subset (large-scale audio dataset with 527 classes used for pre-trained model development). Dataset analysis revealed coverage of relevant workplace sounds (impacts, alarms, human voices) but identified gaps in domain-specific industrial sounds (machinery failures, structural stress), necessitating custom data collection for fine-tuning. These datasets provide baseline training data and establish benchmark comparisons with existing research.

* **Techniques**:

Academic Database Search: Google Scholar, IEEE Xplore, and ArXiv were used to identify state-of-the-art audio classification methods, yielding 5 core papers and 15+ supporting references

Competitive Intelligence: Product websites, case studies, and technical datasheets were analysed to understand existing solutions and identify market gaps

Regulatory Analysis: HSE UK reports, GDPR documentation, and ISO standards were reviewed to ensure legal compliance and safety standard adherence

Public Dataset Collection: Three benchmark audio datasets would be collected and analysed:

UrbanSound8K (urbansounddataset.weebly.com): 8,732 labeled sound excerpts covering urban sounds including glass breaking, sirens, and human activity

ESC-50 (github.com/karolpiczak/ESC-50): 2,000 environmental audio clips across 50 classes, providing diverse sound event examples

AudioSet (research.google.com/audioset): Subset of Google's large-scale audio dataset (2M clips, 527 classes) used as basis for pre-trained models (PANNs, YAMNet)

* **Requirement Specification**

**FUNCTIONAL REQUIREMENTS**

**FR1: Audio Capture and Preprocessing**

Continuously capture audio at 48kHz, 16-bit resolution with 1-second buffering

Support multiple microphone inputs (1-8 channels) with preprocessing (normalization, windowing)

Preprocess audio by normalizing amplitude and applying windowing functions

Handle microphone connection failures with automatic reconnection

Support mono and multi-channel audio configurations

**FR2: Sound Event Classification**

Classify audio into minimum 5 hazard categories: equipment malfunction, falling objects/collision, glass breaking/structural failure, human distress, emergency alarms

Output probability scores for all categories (SoftMax distribution)

Provide confidence scores for top-1 predicted class

Support "Normal/Background" class for non-hazardous ambient sounds

Enable administrators to add new sound classes through retraining

**FR3: Alert Generation and Notification**

Trigger alerts when hazard confidence exceeds configurable threshold (default 80%)

Support multiple alert channels: visual (LED, dashboard), audio (buzzer), network (SMS, email, webhook)

Include in alerts: timestamp, detected class, confidence score, microphone ID, optional audio snippet

Provide alert acknowledgment interface for users

Rate-limit alerts to prevent flooding (max 1 per second per class)

**FR4: Event Logging and Storage**

Log all detected events to local SQLite database with comprehensive metadata

Store timestamp, predicted class, confidence, all class probabilities, microphone ID, alert status

Retain event logs for minimum 30 days with automatic archival

Provide CSV export functionality for external analysis

Optionally record 5-second audio snippets when alerts triggered

**NON-FUNCTIONAL REQUIREMENTS**

**NFR1: Performance**

Achieve end-to-end latency <1 second (95th percentile) from audio capture to alert

Maintain >90% classification accuracy on workplace-specific validation dataset

Achieve false positive rate <10% and true positive rate >85%

Model inference time <500ms on Raspberry Pi 4

Audio capture pipeline latency <100ms

**NFR2: Resource Constraints**

Operate on Raspberry Pi 4 Model B with 4GB RAM

Model file size <50MB (TensorFlow Lite format)

Runtime RAM usage <1GB during continuous operation

CPU utilization <80% average to allow headroom

Storage <8GB for 30 days logs (without audio) or <32GB (with audio)

**NFR3: Privacy and Security**

Process all audio on-device; no raw audio transmitted to external servers

Encrypt audio snippets at rest using AES-256 if recording enabled

Require authentication (bcrypt password hashing) for dashboard access

Comply with GDPR: data minimization, purpose limitation, 30-day auto-deletion, user data export/deletion rights

Implement API rate limiting (max 100 requests/minute per IP)

**NFR4: Reliability and Availability**

Achieve >99% uptime during operational hours

Automatically restart after crashes within 30 seconds

Gracefully handle errors: microphone disconnection (retry every 30s), network outage (queue alerts), database locks (exponential backoff)

Operate continuously for minimum 30 days without manual restart

Continue core monitoring during network outages (edge independence)

**NFR5: Usability and Maintainability**

Complete system installation in <30 minutes for technical users

Initial configuration (microphone, thresholds) in <2 hours

Dashboard accessible from standard browsers without plugins

Provide clear, actionable alerts with plain language and location information

Open-source codebase (MIT License) with >70% test coverage and comprehensive documentation

# Time Schedule

* **Gantt Chart Overview**:

**Figure 1. Gantt chart**

# References / Bibliography

* + - * 1. HSE (2023). Work-related fatal injuries in Great Britain - HSE. [online] Hse.gov.uk. Available at: <https://www.hse.gov.uk/statistics/fatals-overview.htm>. [Accessed 26 Oct. 2025]
        2. Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J. and Wilson, K. (2017). *CNN architectures for large-scale audio classification*. [online] IEEE Xplore. doi: https://doi.org/10.1109/ICASSP.2017.7952132. [Accessed 26 Oct. 2025]
        3. Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W. and Plumbley, M.D. (2020). PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, [online] 28, pp.2880–2894. doi: <https://doi.org/10.1109/taslp.2020.3030497>. [Accessed 26 Oct. 2025]
        4. Elangovan, M., Surrya Prakash, D. and Sasidharan, P. (2021). Monitoring of Workplace Safety Using IoT. Journal of Physics: Conference Series, 2115(1), p.012014. doi: <https://doi.org/10.1088/1742-6596/2115/1/012014>. [Accessed 26 Oct. 2025]
        5. Mu, W., Yin, B., Huang, X., Xu, J. and Du, Z. (2021). Environmental sound classification using temporal-frequency attention based convolutional neural network. Scientific Reports, 11(1). doi: https://doi.org/10.1038/s41598-021-01045-4. [Accessed 30 Oct. 2025]
        6. Tsalera, E., Papadakis, A. and Samarakou, M. (2021). Comparison of Pre-Trained CNNs for Audio Classification Using Transfer Learning. Journal of Sensor and Actuator Networks, 10(4), p.72. doi: https://doi.org/10.3390/jsan10040072. [Accessed 26 Oct. 2025]
        7. Sonitrol.com. (2025). Verified Audio & Video Intrusion | Sonitrol. [online] Available at: https://www.sonitrol.com/solutions/verified-audio-video-intrusion [Accessed 8 Nov. 2025].
        8. March Networks. (2025). AI Series — March Networks. [online] Available at: https://www.marchnetworks.com/products/cameras/ai-series/ [Accessed 8 Nov. 2025].
        9. Health, Safety and Vape Detector | HALO Smart Sensor. (n.d.). Health, Safety and Vape Detector | HALO Smart Sensor. [online] Available at: <https://halodetect.com/>. [Accessed 8 Nov. 2025]
        10. Hse.gov.uk. (2023). Non-fatal injuries at work in Great Britain - HSE. [online] Available at: <https://www.hse.gov.uk/statistics/causinj/overview.htm>. [Accessed 8 Nov. 2025]
        11. Cakir, E., Parascandolo, G., Heittola, T., Huttunen, H. and Virtanen, T. (2017). Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(6), pp.1291–1303. doi:https://doi.org/10.1109/taslp.2017.2690575. [Accessed 8 Nov. 2025]

# Ethics Form