**THESIS** Evaluation of Audio Event Detection Machine Learning Models for Gunshots in Offline Browser Clients

**STUDENT** Alexander Hedre

**SUPERVISORS** Noric Couderc (LTH), Axel Nilsson (Axis Communications), Alex Gustafsson (Axis Communications)

**EXAMINER** Elin A. Topp (LTH)

# Client-side Gunshot Detection: Real-time Machine Learning in Browsers

POPULAR SCIENCE SUMMARY **Alexander Hedre**

Real-time gunshot detection is crucial in surveillance, but current systems rely on cloud servers which raise security concerns. Could a solution, running solely on the client-side, provide the speed and reliability needed to make real-time detection feasible?

In some countries, police officers wear body cameras that send live video to a control center, where operators monitor the footage in real time. Often, a single operator has to watch several video feeds at the same time. In tense situations where a lot happens, the operator can easily become overwhelmed, especially with the audio. Imagine trying to watch three movies at once, the sounds blur together causing a mess. To help with this, people mainly use a cloud service, renting a powerful computer to run an automatic detection tool on the live video and audio to ensure that important events like gunshots are instantly identified. That works, but it also means sending data that is considered sensitive to a third-party client, which can raise security concerns. Instead, we would like a solution that is able to analyze the video stream entirely on the operator's device, preferably in the browser used to display the video.

The goal of this thesis was to develop a system capable of detecting gunshots in real time, using only a browser without relying on cloud-based processing. Research surrounding machine learning in the browser has advanced extensively, however it is still in its infancy and most models today are developed solely for cloud-based computing using Python. The challenge of this thesis was to find and adapt models to solve the task of gunshot detection while ensuring that the throughput of the system was high enough to effectively handle live audio with a high stability. Additionally, to ensure that operators could still use the browser for other tasks, such as rendering the real-time video, it was crucial that the models ran efficiently without causing any interruptions or slowing down those processes.

I adapted two machine learning models, AST and YAMNet, designed to detect over 500 types of audio events, to work in the browser. Since these models were not specialized for gunshots I trained an additional neural network, to interpret their outputs, specifically for gunshot detection.

The models performed well, providing predictions with high accuracy. One surprising result was that the system achieved a higher throughput than a comparable cloud-based solution on the same device, effectively handling real-time audio with greater efficiency. While there were more latency spikes in the system compared to in the cloud-based environment, the system remained stable overall and was able to smoothly run without causing any disruptions to other processes running on the browser simultaneously.

Although this thesis focused on gunshot detection, the method could be adapted for other audio tasks, such as detecting alarms or isolating voices in noisy environments. In conclusion, I have shown that it is possible to run a real-time gunshot detection model in a web browser, offering a viable alternative to cloud-dependent machine learning models.