ABSTRAK

Proses diagnosis kerusakan *Continuously Variable Transmission* (CVT) pada skuter matik saat ini masih dilakukan secara manual dengan membongkar blok transmisi yang memakan waktu dan rentan penilaian subjektif. Penelitian sebelumnya yang menggunakan *Convolutional Neural Network* (CNN) untuk diagnosis suara mesin memiliki kelemahan pada *local receptive field*, sehingga kinerjanya menurun drastis saat menghadapi *noise* mesin pembakaran. Sebagai solusi, penelitian ini mengusulkan arsitektur *Audio Spectrogram Transformer* (AST) berbasis mekanisme *Self-Attention* untuk menangkap konteks frekuensi secara global. Model dilatih menggunakan *Hybrid Acoustic Dataset* (observasi bengkel dan media sosial) yang diseimbangkan melalui *Undersampling Absolute* menjadi 300 sampel per kelas. Audio masukan berdurasi 3 detik diekstraksi menjadi *Mel-Spectrogram*, dikonversi ke skala desibel (dB), dan diproses oleh AST. Sistem ini diintegrasikan ke dalam *Progressive Web App* (PWA) menggunakan API *WebRTC* untuk perekaman langsung. Hasil pengujian menunjukkan model mampu mengklasifikasikan tiga kondisi (Normal, Kerusakan Ringan, dan Kerusakan Parah) dengan Akurasi dan F1-Score sebesar 55%. Pada lingkungan perekaman dunia nyata (*in-the-wild*) dengan *noise* tinggi, sistem ini terbukti potensial diimplementasikan sebagai asisten mekanik virtual berbasis *Condition-Based Maintenance* (CBM) yang objektif guna menjadi *Early Warning System*.

Kata Kunci: *Audio Spectrogram Transformer*, *Continuously Variable Transmission*, *Progressive Web App*, Deteksi Anomali Suara, *Machine Condition Monitoring*.

ABSTRACT

The diagnosis process for Continuously Variable Transmission (CVT) damage in automatic scooters is currently still performed manually by dismantling the transmission block, which is time-consuming and prone to subjective assessment. Previous studies utilizing Convolutional Neural Networks (CNN) for engine sound diagnosis suffer from limitations in their local receptive field, causing their performance to drop drastically when faced with combustion engine noise. As a solution, this study proposes an Audio Spectrogram Transformer (AST) architecture based on the Self-Attention mechanism to capture the global frequency context. The model was trained using a Hybrid Acoustic Dataset (gathered from workshop observations and social media), which was balanced via Absolute Undersampling to 300 samples per class. A 3-second audio input is extracted into a Mel-Spectrogram, converted to a decibel (dB) scale, and processed by the AST. The system is then integrated into a Progressive Web App (PWA) using the WebRTC API for real-time recording. Evaluation results demonstrate that the model is capable of classifying three conditions (Normal, Mild Damage, and Severe Damage) with an Accuracy and F1-Score of 55%. In high-noise, in-the-wild recording environments, this system proves its potential to be implemented as an objective virtual mechanic assistant based on Condition-Based Maintenance (CBM), serving as an Early Warning System.

Keywords: *Audio Spectrogram Transformer, Continuously Variable Transmission, Progressive Web App, Sound Anomaly Detection, Machine Condition Monitoring.*