

## सार

हाल के वर्षों में, न्यूरल नेटवर्क्स (एनएन) ने विभिन्न क्षेत्रों में अनुप्रयोगों में जबरदस्त वृद्धि का अनुभव किया है, जिससे स्वास्थ्य देखभाल, कृषि, निगरानी और स्वायत्त प्रणालियों जैसे उद्योगों में क्रांति आ गई है। बड़े डेटासेट और शक्तिशाली कंप्यूटिंग सिस्टम की उपलब्धता, साथ में टेन्सरफ्लो और पायटोरच जैसी उपयोगकर्ता-अनुकूल लाइब्रेरीज़ ने एनएन को मानव-जैसा प्रदर्शन प्राप्त करने में सक्षम बनाया है। परिणामस्वरूप, एनएन-आधारित एप्लिकेशन चेहरे की पहचान और चैटबॉट से लेकर शॉपिंग अनुशंसाओं और चिकित्सा निदान तक हमारे दैनिक जीवन का एक अभिन्न अंग बन गए हैं।

यह पीएच.डी. थीसिस सीमित कंप्यूटिंग संसाधनों और तंग ऊर्जा बजट वाले एज उपकरणों के लिए न्यूरल नेटवर्क (एनएन) के अनुमान चरण को अनुकूलित करने पर केंद्रित है। स्मार्टफोन और पहनने योग्य डिवाइस जैसे एज डिवाइस अब व्यापक रूप से एनएन अनुप्रयोगों से सुसज्जित हैं, जिससे बेहतर उपयोगकर्ता अनुभव और गोपनीयता सुरक्षा के लिए ऊर्जा-कुशल समाधान की आवश्यकता होती है। अध्ययन में एनएन के तीन वर्ग शामिल हैं: सेल्फ ऑर्गनाइजिंग मैप्स (एसओएम), कन्वेन्शनल न्यूरल नेटवर्क्स (सीएनएन), और रिकरंट न्यूरल नेटवर्क्स (आरएनएन), प्रत्येक वर्ग डेटा पुनः उपयोग और ऊर्जा दक्षता के मामले में अद्वितीय चुनौतियां पेश करता है।

थीसिस न्यूरल नेटवर्क मॉडल को पेश करने से शुरू होती है, जिसमें फीडफॉरवर्ड, रिकरंट और डीप न्यूरल नेटवर्क सहित संरचना और विभिन्न प्रकार के एनएन को दर्शाया गया है। एनएन अनुप्रयोगों के विकास और तैनाती चरणों को भी समझाया गया है, जिसमें एज उपकरणों के लिए अनुमान चरण को अनुकूलित करने के महत्व पर जोर दिया गया है।

सीएनएन के संदर्भ में, अनुसंधान विभाजन और शेड्यूलिंग योजनाओं के माध्यम से डेटा के पुनः उपयोग को अनुकूलित करने पर केंद्रित है। सिस्टम-आर्किटेक्चरल बाधाओं और डेटा आकार पर विचार करते हुए, ऑफ-चिप मेमोरी एक्सेस का अनुमान लगाने के लिए एक विश्लेषणात्मक ढांचा विकसित किया गया है। ढांचा विभिन्न विभाजन और शेड्यूलिंग दृष्टिकोणों की तुलना करने में सक्षम बनाता है, जिससे विभिन्न सीएनएन परतों के लिए ऊर्जा दक्षता और थ्रूपुट में सुधार के लिए इष्टतम समाधान की खोज की सुविधा मिलती है।

आरएनएन के लिए एक नया डेटा पुनः उपयोग दृष्टिकोण प्रस्तावित है, जो पिछले समय-चरण गणनाओं पर निर्भरता के कारण अतिरिक्त चुनौतियां पेश करता है। यह कार्य ऑन-चिप मेमोरी क्षमता के बावजूद, किसी भी दो सन्निहित समय चरणों के लिए बड़े मैट्रिक्स के वजन का कुशलतापूर्वक पुनः उपयोग करता है। लॉन्ग-शॉर्ट टर्म मेमोरी नेटवर्क (एलएसटीएम) एक्सेलेरेटर का एफपीजीए कार्यान्वयन लोकप्रिय एलएसटीएम मॉडल के लिए ऊर्जा दक्षता और थ्रूपुट बढ़ाने में इस दृष्टिकोण की प्रभावशीलता को प्रदर्शित करता है।

एसओएम के लिए, अध्ययन सटीकता और ऊर्जा दक्षता पर परिमाणीकरण तकनीकों के प्रभाव का पता लगाता है। एक कस्टम सेमी-सिस्टोलिक सरणी डिज़ाइन तैयार किया गया है, और विभिन्न बिट रिज़ॉल्यूशन कार्यान्वयन के लिए एनएन सटीकता और ऊर्जा खपत के बीच एक व्यापार-बंद का विश्लेषण किया गया है। अध्ययन एसओएम के लिए अलग-अलग बिट रिज़ॉल्यूशन का उपयोग करने के लाभों और व्यापार-बंदों का विश्लेषण करता है, ऊर्जा-सीमित प्रणालियों को पूरा करता है जहां क्षेत्र, शक्ति और प्रदर्शन महत्वपूर्ण विचार हैं।

अंत में, यह पीएच.डी. थीसिस एज उपकरणों के लिए आधुनिक एनएन के ऊर्जा-कुशल त्वरण में अत्याधुनिक को आगे बढ़ाती है। उपन्यास डेटा पुनः उपयोग तकनीकों और विश्लेषणात्मक ढांचे का प्रस्ताव करके, यह शोध अनुमान चरण में सुधार के लिए मूल्यवान अंतर्दृष्टि प्रदान करता है, इस प्रकार ऊर्जा-सीमित वातावरण में एनएन-आधारित अनुप्रयोगों को व्यापक रूप से अपनाने में योगदान देता है। यह कार्य एनएन अनुकूलन और एज एआई अनुप्रयोगों की सीमाओं को आगे बढ़ाते हुए भविष्य के अनुसंधान के लिए नए रास्ते भी खोलता है।