(12) PATENT APPLICATION PUBLICATION

(19) INDIA

(22) Date of filing of Application :01/07/2025

(21) Application No.202511062876 A

(43) Publication Date : 18/07/2025

(54) Title of the invention : A COMPACT NEURAL PROCESSING UNIT DEVICE FOR EDGE AI DEPLOYMENT

(51) International classification	:G06N0003063000, G06N0003080000, G06F0007544000, G06N0003048000, G06N0003040000
(86) International Application No Filing Date	:NA :NA
(87) International Publication No	: NA
(61) Patent of Addition to Application Number Filing Date	:NA :NA
(62) Divisional to Application Number Filing Date	:NA :NA

(71)**Name of Applicant :**
1)NOIDA INSTITUTE OF ENGINEERING & TECHNOLOGY
Address of Applicant :19, Knowledge Park-II, Institutional Area, Greater Noida – 201306, Uttar Pradesh, India. ---------- ----------
Name of Applicant : NA
Address of Applicant : NA
(72)**Name of Inventor :**
1)NAMITA SHARMA
Address of Applicant :Department of Master of Computer Applications, Noida Institute of Engineering & Technology, Greater Noida. Greater Noida ---------- ----------

(57) Abstract :
A compact neural processing unit device (100) is disclosed, featuring an integrated neural core (102), low-latency memory (104), control processor (106), and multi-interface module (108). The device is optimized for edge AI deployment, enabling real-time inference of deep learning models with minimal power consumption and reduced latency. Supporting quantized and mixed-precision computation, the device ensures compatibility with popular AI frameworks. Its modular architecture and thermal efficiency make it ideal for integration in IoT, mobile, and embedded systems.

No. of Pages : 14 No. of Claims : 5