CLAIMS:

1. A computer-implemented method for secure training of a neural network comprising a first set of layers and a second set of layers, the first set of layers and second set of layers being non-overlapping, comprising:

receiving a label and an encrypted input from a client;

performing homomorphic computations on the encrypted input using the first set of layers to obtain an encrypted intermediate ciphertext;

providing the encrypted intermediate ciphertext to the client;

receiving a non-sensitive decrypted input from the client, wherein the non-sensitive decrypted input corresponds to an output of an activation function applied to a decryption of the encrypted intermediate ciphertext;

performing computations on the non-sensitive decrypted input using the second set of layers to obtain a decrypted output; and

updating weights of the neural network based on the decrypted output.

2. The computer-implemented method of claim 1, wherein receiving the non-sensitive decrypted input comprises receiving a decryption of encrypted gradients derived from the encrypted intermediate ciphertext.

3. The computer-implemented method of claim 1 or 2, wherein performing homomorphic computations on the encrypted input comprises applying a first linear function, utilising weights of the first set of layers, to the encrypted input to produce the encrypted intermediate ciphertext.

4. The computer-implemented method of any one of claims 1 to 3, wherein the activation function includes multiple sign functions or multiple step functions.

5. The method of any one of claims 1 to 4, wherein performing computations on the non-sensitive decrypted input comprises applying an activation function, utilising weights of the second set of layers, to the non-sensitive decrypted input.

6. The method of any one of claims 1 to 5, wherein the neural network comprises at least one further set of layers, the first set of layers, second set of

layers and each further set of layers being non-overlapping, wherein performing computations on the non-sensitive decrypted input comprises sequentially applying the second set of layers, to the non-sensitive decrypted input, and then each further set of layers to an output of the second set of layers or an output of a previous said further set of layers, the method comprising evaluating a loss function over an output of the neural network and the label.

7. The method of any one of claims 1 to 6, wherein each said set of layers comprises a single layer.

8. A computer-implemented method for secure training of a neural network comprising a first set of layers and a second set of layers, the first set of layers and second set of layers being non-overlapping, comprising:

generating, at a client terminal, an encrypted input from a decrypted input;

transmitting the encrypted input and a corresponding label to a server;

performing homomorphic computations, at the server, on the encrypted input using the first set of layers to obtain an encrypted intermediate ciphertext;

providing the encrypted intermediate ciphertext to the client terminal;

generating, at the client terminal, a non-sensitive decrypted input corresponding to an output of an activation function applied to a decryption of the encrypted intermediate ciphertext, and transmitting the non-sensitive decrypted input to the server;

performing, at the server, computations on the non-sensitive decrypted input using the second set of layers to obtain a decrypted output; and

updating weights of the neural network based on the decrypted output.

9. The computer-implemented method of claim 8, wherein the client terminal derives encrypted gradients from the encrypted intermediate ciphertext, and generates decrypted gradients from the encrypted gradients, the non-sensitive decrypted input comprising the decrypted gradients.

10. The computer-implemented method of claim 8 or 9, wherein performing the homomorphic computations on the encrypted input comprises applying a first linear function, utilising weights of the first set of layers, to the encrypted input to produce the encrypted intermediate ciphertext.

11. The computer-implemented method of any one of claims 8 to 10, wherein the activation function includes multiple sign functions or multiple step functions.

12. The computer-implemented method of any one of claims 8 to 11, wherein updating weights of the neural network based on the decrypted output comprises evaluating, at the server, a loss function based on the decrypted output and the label.

13. The computer-implemented method of claim 12, wherein updating weights of the neural network based on the decrypted output comprises evaluating gradients of the loss function and weights of the neural network at both the server and client terminal, and updating the weights based on the evaluation.

14. The computer-implemented method of any one of claims 8 to 13, wherein performing computations on the non-sensitive decrypted input comprises applying an activation function, utilising weights of the second set of layers, to the non-sensitive decrypted input.

15. The computer-implemented method of any one of claims 8 to 14, wherein the neural network comprises at least one further set of layers, the first set of layers, second set of layers and each further set of layers being non-overlapping, wherein performing computations on the non-sensitive decrypted input comprises sequentially applying the second set of layers, to the non-sensitive decrypted input, and then each further set of layers to an output of the second set of layers or an output of a previous said further set of layers, the method comprising evaluating a loss function over an output of the neural network and the label.

16. The computer-implemented method of any one of claims 8 to 15, wherein each said set of layers comprises a single layer.

17. A system for secure training of a neural network comprising a first set of layers and a second set of layers, the first set of layers and second set of layers being non-overlapping, comprising:

at least one memory; and

at least one processor communicatively coupled to the at least one memory and configured to perform the method for secure training of a neural network according to any one of claims 1 to 7.

18. A system for secure training of a neural network comprising a first set of layers and a second set of layers, the first set of layers and second set of layers being non-overlapping, comprising:

a client terminal comprising:

at least one memory; and

at least one processor communicatively coupled to the at least one memory; and

a server comprising:

at least one memory; and

at least one processor communicatively coupled to the at least one memory,

wherein the at least one processor of the client terminal and the at least one processor of the server are configured to perform the method for secure training of a neural network according to any one of claims 8 to 16.