TABLE I
FEATURES USED TO PREDICT THE RELEVANCE BETWEEN CODE COMMITS AND THE VULNERABILITY

Category	Feature	Meaning
Lines of Codes	Added Line Num	The number of lines of code added in the code commit.
	Deleted Line Num	The number of lines of code deleted in the code commit.
	Modified Line Num	The number of lines of code modified in the code commit.
Identity	CVE Num	The number of CVE IDs in the commit message of the code commit.
	Bug Num	The number of Bug IDs in the commit message of the code commit.
	Issue Num	The number of Issue IDs in the commit message of the code commit.
	CVE Match	Whether the code commit mentions the CVE ID in the NVD page.
	Bug Match	Whether the code commit mentions the Bug ID in the NVD page.
	Issue Match	Whether the code commit mentions the Issue ID in the NVD page.
	CWE Match	Whether the code commit mentions the CWE name in the NVD page.
	Vulnerability Type Relevance	The relevance of the vulnerability texts between NVD and the code commit (a three-tuple).
Location	Time Interval	Time interval between the code commit time and the CVE ID assigned time.
	Same Filepath Num	The number of filepaths that exist in both the code commit and the vulnerability description.
	Same Filepath Ratio	Same Filepath Num / The number of filepaths modified by the code commit.
	Unrelated Filepath Num	The number of filepaths that exist in the code commit but not mentioned in the vulnerability description.
	Same File Num	The number of files that exist in both the code commit and the vulnerability description.
	Same File Ratio	Same File Num / The number of files modified by the code commit.
	Unrelated File Num	The number of files that exist in the code commit but not mentioned in the vulnerability description.
	Same Function Num	The number of functions that exist in both the code commit and the vulnerability description.
	Same Function Ratio	Same Function Num / The number of functions modified by the code commit.
	Unrelated Function Num	The number of functions that exist in the code commit but not mentioned in the vulnerability description.
Token	TF-IDF Similarity	The cosine similarity between the TF-IDF vectors of the vulnerability description and the code commit message.
	Same Msg_CWE_Token Num	The number of the same tokens between the code commit message and CWE name.
	Same Msg_CWE_Token Ratio	Same Msg_CWE_Token Num / The number of CWE name tokens.
	Same Msg_Desc_Token Num	The number of the same tokens between the vulnerability description and the code commit message.
	Same Msg_Desc_Token Ratio	Same Msg_Desc_Token Num / The number of tokens in the vulnerability description.
	Max_Desc	The max of the frequencies for all Same Msg_Desc_Token.
	Sum_Desc	The sum of the frequencies for all Same Msg_Desc_Token.
	Mean_Desc	The mean of the frequencies for all Same Msg_Desc_Token.
	Var_Desc	The variance of the frequencies for all Same Msg_Desc_Token.
	Same Code_Desc_Token Num	The number of the same tokens between the vulnerability description and the code change.
	Same Code_Desc_Token Ratio	Same Code_Desc_Token Num / The number of tokens in the vulnerability description.
	Max_Code	The max of the frequencies for all Same Code_Desc_Token.
	Sum_Code	The sum of the frequencies for all Same Code_Desc_Token.
	Mean_Code	The mean of the frequencies for all Same Code_Desc_Token.
	Var_Code	The variance of the frequencies for all Same Code_Desc_Token.